A new treatment planning concept accounting for prompt gamma imaging for proton range verification

Liheng Tian

München 2020

A new treatment planning concept accounting for prompt gamma imaging for proton range verification

Liheng Tian

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Liheng Tian
aus Kunming, China

München, den 9. Juli 2020

Erstgutachter: Prof. Dr. Katia Parodi
Zweitgutachter: Prof. Dr. Marco Riboldi
Tag der mündlichen Prüfung: 27. Juli 2020

Contents

List of Figures

List of Tables

Zusammenfassung

Protenentherapie zur Behandlung von Tumorerkrankungen ist stark abhängig von der Präzision des Protonenstrahls. Die Reichweite von Protonen unterliegt Unsicherheiten, zum Beispiel verursacht durch die Positionierung des Patienten, durch die semi-empirische Umwandlung von Röntgen-CT-Werten in das gewebespezifische Bremsvermögen (relativ zu dem von Wasser), oder durch anatomische Veränderungen. Prompt-Gamma-Bildgebung (PG) eröffnet die Möglichkeit, die Protonenreichweite in vivo zu beobachten, indem Prompt-Gammas detektiert werden, die durch Kernrelaxation innerhalb des Strahlengangs emittiert werden. Studien haben gezeigt, dass die Genauigkeit von PG von verschiedenen Faktoren abhängt, zum Beispiel von der Strahlstatistik, der Gewebeheterogenität, anatomischen Veränderungen und vom PG-Kamera Prototypen. Diese Effekte werden beim konventionellen Protonen-Bestrahlungsplanungsprozess nicht berücksichtigt, weshalb PG-Bildgebung noch nicht vollumfänglich möglich ist. In dieser Dissertation werden diese Effekte untersucht, die sowohl die PG Emission als auch Detektion betreffen. Ein neues Bestrahlungsplanungskonzept wird vorgeschlagen, das die PG-basierende in-vivo Reichweitenverifikation berücksichtigt.

Im ersten Teil dieser Arbeit wird das Monte Carlo (MC) Simulationstoolkit Geant4 verwendet um die Spot-by-Spot-Dosis, die PG Emission und die PG Detektion zu simulieren. Dabei wird eine ideale Scoring-Ebene auerhalb des Targets verwendet (Phasenraumdaten, genannt phsp Daten). Die MC Bestrahlungspläne wurden erstellt mit Hilfe einer Forschungs-Softwareplattform, in der mit MC vorberechnete Nadelstrahlen (PB) mit dem analytischen, Matlab-basierten Bestrahlungsplanungssystem (TPS) CERR kombiniert wurden. Der Effekt der Protonenstrahl-Statistik auf die PG Genauigkeit wird zunächst untersucht, indem lateral integrierte Profile der phsp Daten des selben PBs in verschiedenen Szenarien verglichen werden. Die Ergebnisse zeigen, dass mindestens 1.35×10^8 Protonen pro PB nötig sind, um eine PG Genauigkeit von 1mm zu erreichen. Zweitens wird die Konformität zwischen der Dosis und der PG Emission (PG Dosiskorrelation) untersucht und quantifiziert indem der 80%-distale Abfall und der Abfall der PG Emissionsprofile verglichen wird mithilfe eines Machine Learning Mustererkennungsalgorithmus. Beim ursprünglichen Bestrahlungsplan konnten nur wenige PBs aus dem Optimierungsprozess eine ausreichende Statistik erreichen und nicht alle PBs hatten eine gute PG-Dosiskorrelation, insbesondere solche mit dem Bragg-Peak in der Nähe von Gewebeheterogenitäten, z.B. Luftkavitäten. In dieser Arbeit wird ein neues Bestrahlungsplanungskonzept vorgeschlagen, das PG Bildgebung berücksichtigt und wenige PBs verstärkt, basierend auf der

PG-Dosiskorrelation, sodass sie eine ausreichende Statistik haben, während die Qualität der Bestrahlung unverändert bleibt. Die neuen Bestrahlungspläne wurden verglichen mit den ursprünglichen Bestrahlungsplänen basierund auf einer MC Simulation auf der selben Computertomographie (CT) für 3 Kopf-Hals-Tumoren Patienten. Es wird gezeigt, dass die neuen re-optimierten Bestrahlungspläne äquivalent sind zu den ursprünglichen bezüglich der Dosisverteilung, der Verteilung des Dosis-gemittelten linearen Energietransfers (LET) und der Robustheit gegenüber Verschiebungen/Reichweitenveränderungen, und gleichzeitig die statistischen Bedingungen für ein zuverlässiges PG-Monitoring der ausgewählten PBs erfüllen.
Der zweite Teil dieser Arbeit beschäftigt sich mit dem Effekt von inter-fraktionalen anatomischen Veränderungen auf die Robustheit der PG-Dosiskorrelation. Dies wurde erreicht durch eine 2D distale Oberfläche (Dosisoberfläche), die von einer 3D MC Dosisverteilung der entsprechenden PBs abgeleitet wurde. Die Dosisoberfläche wurde als Indikator für PBs genutzt, die stärker abhängig von anatomischen Veränderungen waren. Mit Hilfe von CERR und Geant4, wurden die neuen re-optimisierten Bestrahlungspläne basierend auf einer Kombination der Dosisoberfläche und PG-Dosiskorrelations-Indikatoren vergleichen mit den ursprünglichen Bestrahlungsplänen auf den CTs von 1 Kopf-Hals-Tumor und 1 Prostata-Karzinom Patienten zu 3 verschiedenen Zeitpunkten. Die Verschiebung der Dosis und PG-Emissionsprofile auf verschiedenen CTs und mit den selben PBs wurden verglichen um zu entscheiden ob ein gegebener PB für PG Bildgebung verlässlich ist. Es wurde gezeigt, dass die re-optimisierten und ursprünglichen Bestrahlungspläne vergleichbar sind im Sinne der Dosisverteilung und der Verteilung des Dosis-gemittelten LET über alle CTs, während die Verschiebungsunterschiede zwischen der Dosis und PG-Emission der vorgeschlagenen PBs innerhalb von 1 mm waren.
Der letzte Teil dieser Arbeit beschäftigt sich mit der realistischen Antwort einer Knife-edge Slit PG Kamera, welche ein Kandidat ist für die erste klinische Reichweitenbestimmung von PB scanning durch PG. Die Dosisverteilung und die PG Emission und Detektion wurden simuliert mit REGistration Graphical User Interface (REGGUI) für die selben PBs und die selben CTs (nur 2 Zeitpunkte wurden genutzt) aus dem zweiten Teil. Da diese Art von PG-Kamera-Prototyp die korrekte PG Information zur Verfügung stellen soll für PBs deren PG Abfall zentral im Blickfeld (FOV) der Kamera ist, wurde ein zweiter Indikator untersucht basierend darauf, ob ein PB zentral im FOV liegt oder nicht. Dieser Indikator wurde zusammen mit dem Indikator für PG-Dosiskorrelation und der Dosisoberfläche untersucht. Die Dosis und PG Emission generiert durch REGGUI wurden zunächst verglichen mit denen erstellt durch Geant4. Danach wurde das PG Detektionsprofil unter verschiedenen Statistiken des selben PBs verglichen, um den Einfluss der Statistik auf die PG Detektion mittels dieser PG Kamera zu untersuchen. Eine Genauigkeit von 0.8 mm konnte erreicht werden mit 2.00×10^8 Protonen. Als nächstes wurden Verschiebungen der Dosis und PG-Emission verglichen und die Ergebnisse sind konsistent mit den vorhergehenden Schlüssen. Danach wurden die PBs-of-Interest identifiziert basierend auf allen drei Indikatoren und deren Dosis- und PG-Detektionsprofile auf verschiedenen CTs wurden in diversen Szenarien verglichen. Abgesehen von zwei Ausnahmen, die durch unvorhersehbare und vergleichsweise groe anatomische Veränderungen hervorgerufen

wurden, sind alle PG Signale der meisten vorgeschlagenen PBs verlässlich zum Monitoring des Dosisabfalls, was bedeutet, dass der Verschiebungsunterschied zwischen der Dosis und dem detektierten PG-Signal unter 2 mm ist. Im Gegensatz dazu wurde ein Verschiebungsunterschied von bis zu 9.1 mm beobachtet für die abgelehnten PBs. Aus diesem Grund sind diese Indikatoren sehr wertvoll zur Identifikation von verlässlichen PBs für PG-basierte Protonen-Reichweitenverifikation.
Der Ansatz der PB-Verstärkung wird zudem verglichen mit anderen alternativen Ansätzen der PB Aggregation, welche naheliegende PBs aggregieren um eine höhere Statistik zu erhalten, für alle vorgeschlagenen Indikatoren. Beide Ansätze sind vergleichbar bezüglich der Reichweitenverifikation, jedoch zeigt die PB-Verstärkung Vorteile im Sinne einer garantierten Statistik, einer verbesserten lateralen Auflösung und einer verringerten Reichweitenvermischung.
Mit den Ergebnissen in dieser Dissertation wurde zum ersten Mal eine PG-basierte in vivo Protonen-Reichweitenverifikationsmethode integriert in den Protonen-Bestrahlungsplanungsprozess, was möglicherweise zu zukünftigen Verbesserungen der präzisen Protonen-Bestrahlung beitragen könnte.

Abstract

Proton therapy for cancer treatment is highly affected by the accuracy of the proton beam delivery. However, proton range has uncertainties caused by e.g. patient positioning along with the semi-empirical conversion of the X-ray CT numbers into tissue stopping power ratio (relative to water) and anatomical changes. Prompt gamma (PG) imaging provides the possibility to monitor the proton range in vivo by detecting the prompt gammas emitted by nuclei de-excitation in the beam path. Studies point out that the accuracy of PG is affected by factors e.g. the beam statistics, tissue heterogeneity, anatomical changes and the PG camera prototype. These effects are not considered in the conventional proton therapy treatment planning process thus the PG imaging technique is not fully supported. In this thesis, these effects are investigated at both the PG emission and detection level. A new treatment planning concept accounting for PG-based in-vivo proton range verification is proposed.

In the first part of this work, the Monte Carlo (MC) simulation toolkit Geant4 is employed to simulate the spot-by-spot dose, PG emission and PG detection using an ideal scoring plane outside the target (phase space data, called phsp data). The MC treatment plans are made using a research computational platform, combining MC pre-calculated pencil beams (PB) with the analytical Matlab-based treatment planning system (TPS) engine CERR. The effect of the proton beam statistics to the PG precision is firstly investigated by comparing the laterally integrated phsp data profiles of the same PB under different statistics scenarios. The result shows that at least 1.35×10^8 protons per PB are needed to achieve a PG precision of 1 mm. Secondly, the conformity between the dose and PG emission (PG-dose correlation) has been investigated and quantified by comparing the 80% distal dose falloff and the falloff of the PG emission profiles, which is identified using a machine learning based pattern recognition algorithm. In the initial treatment plan, only a few PBs resulting from the optimization process can reach the wanted statistics and not all PBs have good PG-dose correlation, especially those with their Bragg peak near tissue heterogeneity, e.g. an air cavity. In this work, a new treatment planning concept accounting for PG imaging was proposed to boost a few PBs, selected based on the quantification of PG-dose correlation, above the required statistics while maintaining the treatment quality. The new treatment plans were compared to the initial treatment plans using MC simulation on the same computed tomography (CT) for 3 head and neck cancer patients. The new re-optimized treatment plans are shown to be equivalent to the initial ones in terms of dose distribution, dose averaged linear energy transfer (LET) distribution and robustness

with respect to translational changes/range variations, while fulfilling the set statistical conditions for reliable PG monitoring for the selected PBs.
The second part of this work is dedicated to investigate the effect of the inter-fractional anatomical changes on the robustness of PG-dose correlation. This is accomplished by using a 2D distal surface (dose surface) derived from 3D MC dose distribution of the corresponding PBs. The dose surface could be used as indicator of PBs more sensitive to anatomical changes. Using CERR and Geant4, the new re-optimized treatment plans based on the combination of the dose surface and PG-dose correlation indicators are compared to the initial treatment plans on the CTs of 1 head and neck and 1 prostate cancer patients at 3 different time points. The shift of the dose and PG emission profiles on different CTs of the same PB are compared to evaluate if a given PB is reliable or not, in terms of PG imaging. The re-optimized and the initial treatment plans are shown to be comparable in terms of dose distribution and dose-averaged LET distribution on all CTs, while the shift differences between the dose and PG emission of the recommended PBs are within 1 mm.
The last part of this work focuses on the realistic response of a knife-edge slit PG camera, which is under investigation in the first clinical PG based proton range verification research for PB scanning. The dose distribution as well as the PG emission and detection are simulated using REGistration Graphical User Interface (REGGUI) for the same PBs and same patients CTs (only 2 time points are used) employed in the second part. As this kind of PG camera prototype is supposed to provide correct PG information for PBs whose PG falloffs are centered in the field of view (FOV) of the camera, another proposed indicator, identifying whether a given PB is centered in the FOV or not, is investigated together with the PG-dose correlation and dose surface indicators. The REGGUI generated dose and PG emission are firstly compared to those generated by Geant4. Secondly, the PG detection profile under different statistics for the same PB are compared to investigate the effect of the statistics to the PG detection using this PG camera. A precision of 0.8 mm can be achieved using 2.00×10^8 protons. Next, the shifts of the dose and PG emission are compared and the result is shown to be consistent to the previous findings. Then, the PBs of interest are identified by all the three indicators and their dose and PG detection profiles on different CTs are compared accounting for different scenarios. Besides two exceptions caused by unpredictable and comparably large anatomical changes, the PG signals of most of the PBs recommended are proven to be reliable for dose falloff monitoring, i.e. the shifts difference between the dose and detected PG signal are within 2 mm. In contrast, a shifts difference up to 9.1 mm has been observed for the counter-indicated ones. Hence, these indicators are valuable to be used to identify and recommend reliable PBs for PG based proton range verification.
The PB boosting approach is compared to the other alternative approach of PB aggregation, which aggregates nearby PBs to achieve high statistics, for all the indicators proposed. Both approaches are comparable in terms of range verification but the PB boosting shows advantages in terms of guaranteed statistics, improved lateral resolution and reduced range mixing.
With the results of this thesis, the PG based in-vivo proton range verification method is

integrated in the proton treatment planning process for the first time, potentially contributing to a future improvement of precise proton delivery.

Chapter 1

Introduction and motivation

1.1 Introduction of proton therapy

Cancer is a group of diseases involving abnormal cell growth with the potential invasion to adjoining parts of the body. The mechanisms stem from inherited genetics and genetic mutations caused by environment (e.g. pollution) and lifestyle (e.g. tobacco and alcohol). According to (Ferlay et al. 2019) there were an estimated 18.1 million (95% confidence interval: 17.5-18.7 million) new cancer cases and 9.6 million (95% confidence interval: 9.3-9.8 million) deaths from cancer worldwide in 2018.
Treatment of cancer encompass one or more modalities among e.g. surgery, chemotherapy, immunotherapy and radiotherapy (RT). Radiotherapy plays an essential role in cancer management with around 50% patients receiving RT at a certain point of their treatment (Baumann et al. 2016).
The earliest roots of the medical application of radiation go back to the discovery of X-rays in 1895 by Wilhelm Rontgen, which was then quickly followed by the discovery of natural radioactivity and the isolation of radium in 1898 by Marie Curie. Though the hazards of radiation exposure were initially not fully understood and little protection was used until the 1920s (Matanoski et al. 2001), radium was believed to have wide curative powers and radiotherapy was applied to many diseases.
In the following period, radiotherapy was continuously advanced to treat patients affected by deep seated tumors, to overcome the limited penetration depth of X-rays from X-ray tubes or alpha radiation from Radium. Thus, since the 1950s RT with external beams (teletherapy) has evolved to higher ($\geqslant$ 1 MeV) photon energies from radioactive sources (e.g, ^{60}Co) or produced with accelerators, to deliver the treatment starting from an external radiation source at a certain distance from the patient surface. Additionally, radioactive sources of shorter range radiation (e.g., low energy photons and electrons) can be used for internal/interstitial therapy like radionuclide therapy and brachytherapy, where the radiation source is placed at contact with or inside the patient targeting the organ with a tumor.
Nowadays conventional radiotherapy of deep-seated targets typically uses high energy (6-

20 MeV) X-rays generated via Bremsstrahlung from electron linear accelerators (LINACs). Modern photon therapy includes possibilities to rotate (even continuously) the radiation beam around the patient with intensity modulation. Despite these technological advances, including also forms of image guidance to monitor the patient anatomy at the treatment site, photons are limited by the physics of their depth dose (average energy absorbed per unit mass) distribution, resulting in a continuous exponential attenuation after an initial build-up maximum at a shallow (up to few centimeters) depth. To overcome these limitations, heavy charged particles like protons can be regarded as an advanced form of RT, which is meanwhile widely developed and used in state-of-the-art centers. It was in 1946 that Robert Wilson suggested that energetic protons could be an effective RT treatment method (Wilson 1946). In his seminal work, he anticipated the therapeutic benefits of the increased energy deposition of protons and heavier ions at the end of their path, owing to the earlier findings of (Bragg & Kleeman 1904) with the ionization of alpha particles from Ra in air, hence being referred to as Bragg peak. In 1954 the first proton therapy treatment was performed with particle accelerators for physics research in Lawrence Berkeley Laboratory (Tobias et al. 1958). Later in 1989, the first hospital-based proton therapy center, which was a low energy cyclotron center for ocular tumors, opened at the Clatterbridge center for oncology in the UK, followed in 1990 by the dedicated synchrotron based center for deep seated tumours at the Loma Linda university medical center in Loma Linda, US. The creation of a dedicated hospital-based proton facility was the milestone motivating the expansion of facilities and patient treatments around the world, which gathered the attention of the radiation oncology community (Das & Paganetti 2015). By November 2019 there are about 91 active facilities providing treatments using protons and meanwhile 13 with carbon ions, some of which are combined facilities for both carbons and protons, all over the world (PTCOG 2020).

As already recognized in the seminal work of Robert Willson, one of the most important advantages of using protons (or heavier ions) for radiotherapy is that the energy deposition of energetic ($\geqslant 1$ MeV/u) heavy charged particles is inversely proportional to their squared velocity, causing a dose deposition maximum, so-called Bragg peak, at the end of the beam penetration depth, i.e. the range. By changing the initial beam energy and modulating the proton fluence it is possible to overlap multiple Bragg peaks, resulting in a so-called spread-out Bragg peak (SOBP) for good dose coverage to the target volume already with a single beam direction (figure 1.1). More beam portals including fluence modulation can be then used to further enhance target dose and spread the entrance dose over larger volumes as done with photons, but due to the superior physical selectivity only 2-3 proton treatment fields are typically sufficient compared to modern intensity-modulated radiotherapy (IMRT) (Bortfeld 2006) and volumetric modulated arc therapy (VMAT) (Teoh et al. 2011), with considerably reduced normal tissue exposure.

The clinical exploitation of the protons Bragg peak for radiotherapy highly depends on the accuracy of the calculation of the proton range so that the planned dose is delivered in the target while the dose delivered in the healthy tissue is minimized. It can be imagined that high dose might be delivered in the healthy tissue while the target volume would not receive the prescribed dose if the proton range is different from the prediction.

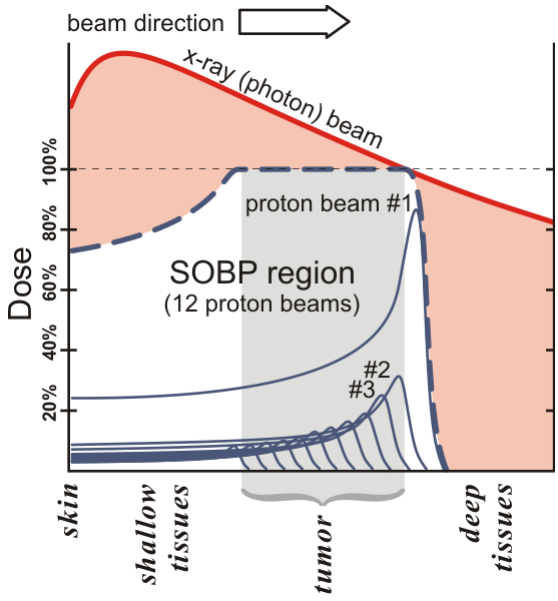


Figure 1.1: Depth dose distribution in water for a photon beam (red), mono-energetic proton beams (solid blue) and their resulting SOBP (dash blue). The photon dose shows a build-up maximum close to the entrance and afterwards decreases with the penetration depth. For protons, the dose increases with depth and reaches a narrow maximum, called Bragg peak, near the end of the profile in correspondence of the beam range. By overlapping multiple proton beams, the SOBP is created to cover the extended tumour target volume with still acceptable entrance dose proximal to and no exit dose distal to the tumour (Levin et al. 2005).

In practice, the proton beam range exhibits non-negligible (typically up to 3%) uncertainties in the patient model used for treatment planning, due to e.g. the semi-empirical conversion of the X-ray computed tomography (CT) numbers into tissue stopping power ratio (relative to water), along with patient positioning errors and anatomical changes in the typically fractionated course of therapy over several weeks. These factors will be further discussed in chapter 2. To mitigate the deleterious effect of the proton range uncertainties, practitioners typically introduce generous safety margins around the target volume and prefer conservative beam directions avoiding to stop the beam in front of radiosensitive targets, thereby not fully exploiting the sharp dose gradients enabled by proton beams. To overcome such limitations, researchers have proposed and investigated over the last two decades various methods to measure and assess the proton range in-vivo, e.g. via positron emission tomography (PET), which detects the pairs of annihilation gamma rays following the β^+ nuclear decay of irradiation induced positron emitters (Kubota 2001).
Another promising approach is prompt gamma imaging (PG), which is the main topic of this thesis work. PG detection for in vivo proton range verification was firstly proposed by (Stichelbaut & Jongen 2003) and then experimentally demonstrated in first phantom studies by (Min et al. 2006) and in first clinical investigations by (Richter et al. 2016, Xie

et al. 2017). Inelastic interactions between the incoming protons and the tissue nuclei occur along the beam path until 2-3 mm before the Bragg peak (depending on the energy threshold of the involved nuclear reactions), thus the PG emission has been found to be well correlated with the distal dose falloff (Knopf & Lomax 2013). Besides, PG is emitted within less than nanoseconds after the nuclear interaction, making a real-time proton range verification possible.

1.2 Motivation of this work

Several studies show that the precision of the PG range verification depends on the statistics (i.e., number of particles) of the incident proton beam (Xie et al. 2017, Draeger et al. 2018). With current detector systems, more than 10^8 protons are needed to generate smooth PG profiles. Furthermore, the accuracy of the PG-based range retrieval could be affected by tissue heterogeneities, potentially spoiling the correlation of the PG signal to the dose delivery (Janssen et al. 2014, Priegnitz et al. 2015, Schmid et al. 2015). Those effects are not considered in traditional proton therapy treatment plans (TP, see chapter 2) as the in vivo proton range verification is not taken into account at the stage of planning. The goal of this thesis is to investigate and quantify the effect of statistics, tissue heterogeneities as well as the response of a real PG camera (Xie et al. 2017) on the PG range verification accuracy and propose a new treatment planning concept accounting for those effects, so that the statistics of few selected and so-called PG-friendly pencil beams (PBs) fulfills the requirement to enable a more reliable treatment monitoring.
This thesis is organized as follows:
Chapter 2 reviews the fundamental aspects of the physics of proton therapy as well as the sources of the proton range uncertainties. A brief review of the start-of-the-art in-vivo proton range verification methods, i.e. PET, ionacoustics and PG, is presented.
Chapter 3 then introduces different prototypes for PG imaging and addresses the effect of beam statistics and tissue heterogeneities on the conformity between the PG signal and dose.
The materials and methods used in this study are described in chapter 4, including the Monte Carlo simulation toolkit Geant4 (Agostinelli et al. 2003) extensively used for particle transport and interaction, the treatment planning system CERR (Deasy et al. 2003, Resch et al. 2017), the REGGUI platform (Janssens et al. 2009, Kurz et al. 2015) for analytical modeling of dose and PG emission along with the PG camera detector response, and basic concepts of machine learning employed for pattern recognition of the PG signal falloff in the subsequent chapters.
Chapter 5 presents the original investigations of this work based on dose distributions and PG emission data generated by detailed Monte Carlo simulations using Geant4 on the basis of clinical patient CT datasets obtained from the LMU University Hospital. The first part of this study investigated and quantified the correlation between the profiles of the dose and PG (PG-dose correlation) by comparing their falloff distributions. A new TP

concept was then proposed to boost a few PBs (based on the PG-dose correlation) above the needed statistics, and tested using the research treatment planning engine CERR. The re-optimized TP using such a PB boosting method is compared to the initial TP as well as to the alternative PB aggregation approach introduced in the literature to overcome PG statistics limitations without modifying the planning process (Tian et al. 2018, Tian et al. 2020). The second part of this work focuses on the robustness of the PG-dose correlation in presence of anatomical changes. The effect of the tissue heterogeneity on the robustness of the PG-dose correlation is quantified by using a 2D distal surface derived from 3D dose distributions of the PB. This new indicator is combined with the PG-dose correlation. The new TP and new indicators are evaluated using the same approaches as those used in the first part of the work, however considering patient CT data acquired at different time points to include the effect of anatomical changes.
Chapter 6 studies how the previous results at emission are influenced by the response of a real PG camera, for which the measurement is found to be affected by the relative position of the PG falloff within the camera field of view. By employing a camera positioning indicator, the PB selection indicators introduced in chapter 5 at emission are further expanded to account for the detection process. For this study, the REGGUI platform is used, thus augmenting the application of the new proposed approach with analytically generated PG data at emission and detection, the latter including (as in real measurements) or not (as in ideal analytical calculations) effects of counting statistics. Using the same clinical datasets of chapter 5 with 2 CTs per patient, one ideal and two realistic PG detection scenarios are compared and discussed for the methods of PB boosting proposed in this work and the PB aggregation proposed in the literature. In the last chapter, a summary and outlook of this PhD thesis work is presente

Chapter 2

Theory background

2.1 Physics of proton interactions in matter

2.1.1 Energy loss rate, linear energy transfer, dose and proton range

The total stopping power of protons (shown in figure 2.1) consists of electronic, nuclear and radiative stopping power. For proton therapy, the radiative stopping power is negligible and the nuclear one is only relevant below ∼10 keV/u, hence also negligible. The electronic stopping power is described by the well-known Bethe-Bloch equation (valid only for energies higher than 1 MeV/u):

$$\frac{S_{el}}{\rho} = -4\pi N_A {r_e}^2 m_e c^2 \frac{Z}{A}\frac{z^2}{\beta^2}[ln\frac{2m_e c^2\gamma^2\beta^2}{I} - \beta^2 - \frac{\delta}{2} - \frac{C}{Z}] \tag{2.1}$$

in which S_{el} is the linear electronic stopping power, which describes the average rate at which protons lose energy dE per unit path length dx due to inelastic interactions with the atomic electrons:

$$S_{el} = \frac{dE}{dx}[MeVcm^{-1}] \tag{2.2}$$

and where ρ is the density of the material, N_A is the Avogadro number, r_e is the classical electron radius, m_e is the electron mass, z is the charge of the particle (z=1 in the case of proton), Z is the atomic number of the medium, A is the atomic mass number, c is the speed of light, $\beta = u/c$ where u is the velocity of the incident particle, $\gamma = (1-\beta^2)^{-1/2}$ is the Lorentz factor, I is the mean excitation energy of the medium and C is the shell correction term.

When the protons slow down, i.e. β is comparable to the velocity of the orbital electrons, the Bethe-Bloch equation is no longer valid to predict the stopping power. In this region, so-called Lindhard region, the energy loss is proportional to the β, and can be described by the model of Anderson and Ziegler (Ziegler et al. 1985).

The *dose* (Gy) can be defined by the average energy dE (J) absorbed per unit mass dm

(kg):

$$dose = \frac{dE}{dm} \tag{2.3}$$

For heavy charged particles and under the assumption of secondary electrons equilibrium, this can be linked to the particle fluence spectrum times the mass stopping power (i.e. stopping power divided by the density), integrated over the entire energy range:

$$dose = \sum_i \int \frac{\Phi_{E,i}}{\rho}(\frac{dE_i}{dx})dE \tag{2.4}$$

where the $\Phi_{E,i} = d\Phi_i/dE$ is the fluence spectrum of all primary and secondary i-th heavy charged particles (mostly protons).
The linear energy transfer (LET) (L_Δ, $keV/\mu m$), depending on the material as well as the type and energy of the incident charged particle, is given by:

$$L_\Delta = \frac{dE_\Delta}{dx} \tag{2.5}$$

where dE_Δ is the mean energy lost due to electronic interactions in traversing the distance dx, minus the kinetic energies in excess of Δ of all electrons released by the incident particle. The LET L_Δ can also be calculated by: $L_\Delta = S_{el} - (dE_{ke,\Delta}/dx)$ in which the $dE_{ke,\Delta}$ is the mean sum of the kinetic energies, larger than Δ, of all the electrons released by the incident charged particle in traversing a distance of dx (ICRU90 2016). When the Δ is set to infinity, the LET would be equivalent to S_{el} (ICRU90 2016) and be regarded as an indicator of radiation quality which affects the relative biological effectiveness (RBE) of the radiation (Hall 2000, Abolfath et al. 2017). RBE is defined by:

$$RBE = \frac{D_i}{D_\gamma} \tag{2.6}$$

where D_i and D_γ are the absorbed dose of the incident particle beam and photon radiation that causes the same amount of biological damage, respectively (Laramore et al. 2013).
As shown in simulation (figure 2.1 and 2.2), the stopping power of protons increases dramatically with decreasing energy until a maximum value and then decreases again in the very low-energy region where the Bethe-Bloch formula ceases its validity, thus being responsible for the formation of the Bragg peak toward the maximum proton penetration depth (cf. figure 2.2 left). Consequently, high energy is delivered at the Bragg peak where the LET also reaches high value (figure 2.2 right), enabling proton beams to deliver high dose with comparably high RBE (Hall 2000) in the desired region while leaving low dose in the beam path, i.e. healthy tissue in the case of radiotherapy. The average length traveled by protons slowing down from the initial kinetic energy E_0 to 0 is described by the so called continuous slowing down approximation range (R_{CSDA}), defined by:

$$R_{CSDA} = \int_{E_0}^{0} -\frac{dE}{S(E)} \tag{2.7}$$

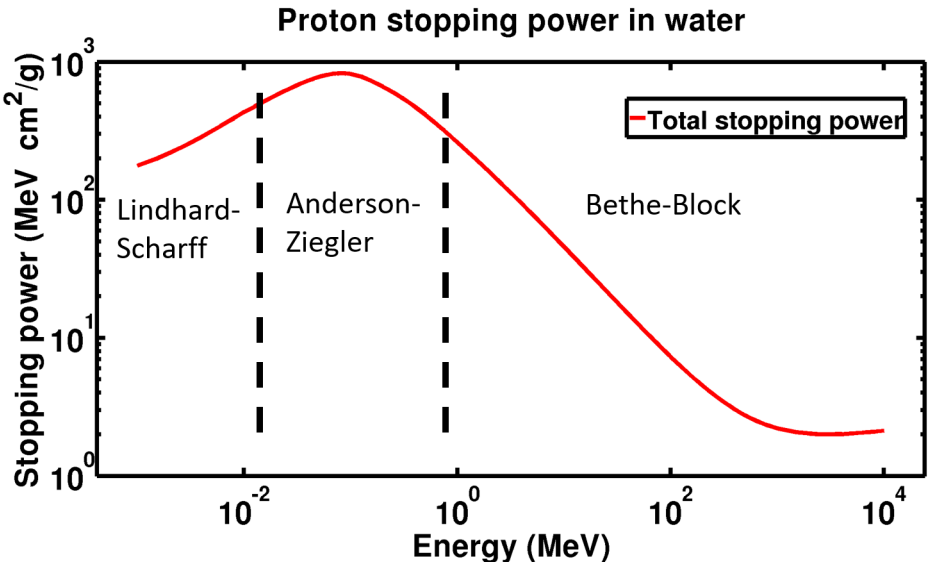


Figure 2.1: Stopping power of protons in water as a function of the kinetic energy made using NIST data, last updated in July 2017 (Berger et al. 2017).

where $S(E)$ is the stopping power calculated as a function of the kinetic energy E. However, due to scattering effects (cf. section 2.1.2), the actual mean depth penetrated by protons in the incident direction, called projected range R_{mf}, is shorter than R_{CSDA}. This is described by a ratio called detour factor R_{mf}/R_{CSDA}.
While penetrating in matter the incident protons spread in the position and momentum phase space due to the stochastic nature of the energy loss process, of which the stopping power only represents the mean value. This spread causes a small variation of the range of individual particles around the mean range value, so called range straggling, which increases with increasing penetration depth (cf. figure 2.2). In practical applications in

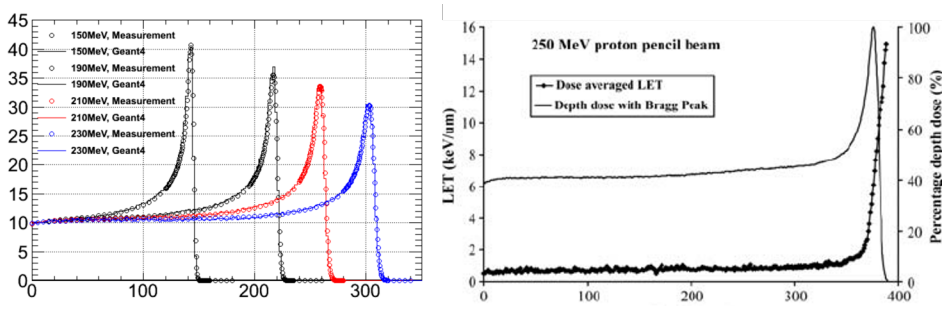


Figure 2.2: Left: Bragg peak of protons in water (depth given in millimeters) for different incident energies based on simulation and measurement data (Ivanchenko et al. 2014). Right: dose averaged LET (LET_d) distribution as well as the corresponding Bragg peak of a 250 MeV proton beam (Chen & Ahmad 2011).

proton therapy, as well as in this work, the proton range is defined by the 80% distal falloff of the laterally integrated depth dose profile, which is found to be less sensitive to such dependencies as range straggling and initial beam momentum spread.

2.1.2 Coulomb scattering

Protons are scattered by various Coulomb interactions with electrons and nuclei in the beam path. The global effect is called multiple Coulomb scattering (MCS). Due to the mass ratio, the Coulomb interaction with nuclei is the dominant scattering mechanism. This effect in a target can be described by the Moliere theory:

$$f(\theta) = \frac{1}{4\pi {\theta_M}^2} \sum_k \frac{f^k(\theta')}{B^k} \tag{2.8}$$

where $f(\theta)$ is the probability distribution function of the scattering angle θ, θ_M is the characteristic MCS angle, $\theta' = \theta/\sqrt{2}\theta_M$ and B is the logarithm of the effective number of collisions in the target. For protons, the lateral scattering increases for thick targets and target materials with high atomic number.
According to the central limit theorem, the probability distribution of the deflection angle of the incident proton in a thick material is nearly Gaussian as the result of numerous small random deflections with a width (standard deviation) given by (Highland 1975):

$$\sigma_\theta = \frac{17.5MeV}{\beta pc}\sqrt{\frac{L}{L_{rad}}}[1 + 0.038ln(\frac{L}{L_{rad}})] \tag{2.9}$$

where L is the total mass thickness, $L_{rad} \sim Z^{-2}$ is the radiation length. The lateral multiple Coulomb scattering increases for thick targets and high Z materials while decreases with increasing incident energy according to the factor of βpc.
The analytical calculation of the MCS effect on the dose distribution is difficult, hence this is typically addressed using Monte Carlo (MC) simulations or empirical parameterizations in treatment planning systems, which also account for the additional low-dose tails resulting from nuclear interactions (cf. section 2.1.3), as described e.g. in (Bellinzona et al. 2015).

2.1.3 Nuclear interactions

Incident protons can also interact with atomic nuclei via the nuclear force, resulting in a reduction of the primary fluence. The target can break up, be excited, or yield a particle transfer reaction. The following function can be used to describe the probability that an incident proton is not having a nuclear interaction with the target atoms:

$$P(x) = e^{-\frac{x}{\lambda_{mfp}}} \tag{2.10}$$

where the λ_{mfp} is the so-called mean free path $\lambda_{mfp} = A/(N_A\sigma\rho)$, A is the atomic mass number, σ is the cross section of nuclear interaction and ρ is the density. An example of

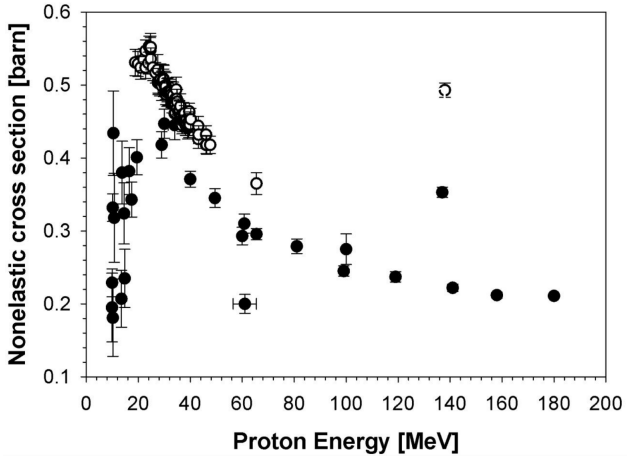


Figure 2.3: Nonelastic nuclear cross section as a function of proton energy. The closed and open circles describe the interaction with the carbon and oxygen target nuclei, respectively (Paganetti 2018).

dependence of the nonelastic nuclear cross section on typical constituents of the human body is shown in figure 2.3.
In the case of proton therapy, the reaction of an incident proton hitting an atomic nucleus in the beam path can be generally described by the following stages:

Intra-nuclear cascade

Assuming that the incident particle interacts with quasi-free nucleons within the nucleus, this model is implemented by (Bertini et al. 1974) for the dynamic part of the reaction. Protons and neutrons have momentum and binding energy following the Fermi gas model and the semi-empirical mass formula (Battistoni et al. 2016). According to the total cross section and the nucleon density, the path length for the projectile can be computed. The type and momentum of the striking particle and the target are determined when the nuclear interaction occurs. In the final state particles are treated as the primary and transported through the nucleus, producing the cascade when the assertion of the Pauli Exclusion Principle and the incident energy is higher than the threshold (Ferrari & Sala 1998). The secondary particles can scatter or escape and through the coalescence mechanism not only protons and neutrons but also light nuclear fragments can be emitted, depending on the energy. Hence, protons, neutrons and other nuclear fragments can be generated (Ribansky et al. 1973, Blann 1983). There are also other available models, e.g. the binary cascade model (Bertini et al. 1974) and the quantum molecular dynamics model (Sorge et al. 1989), which is the most comprehensive hadronic inelastic model not simulating nucleon-nucleon collision but a more collective effect.

Pre-equilibrium

When the energies of the particles of the cascade are lower than the energy threshold of a few tens of MeV but the nucleus is not in thermal equilibrium, the interactions are modeled using a semi-classical approach where nuclear collisions originate holes in the Fermi sea that represent the excited nucleons (Griffin 1966). Particles are emitted under different probabilities for each step. The residual nucleus stays in equilibrium with some excitation energy shared among the remaining nucleons (Battistoni et al. 2016).

De-excitation

Different nuclei then de-excite in different ways, depending on the mass. For heavy nuclei ($Z \geqslant 65$), not considered in proton therapy, the main mechanism is the fission of the excited nucleus into two fragments. For other nuclei, light particles, which have kinetic energies of a few MeV, would be emitted from the excited nuclei, according to the nuclear evaporation approach (Weisskopf 1937). In the so-called Fermi-breakup model proposed by (Fermi 1950), the excitation energy may be above the binding energy of some fragmentation channels, leading to the split of the nucleus into smaller fragments. In the end, gamma rays are released from the remaining excited nuclei, providing the possibility to monitor the range of the incident protons (Stichelbaut & Jongen 2003, Min et al. 2006).

The absorbed depth dose curve is affected by the nuclear interactions. There is a considerable energy and tissue dependent reduction in the primary proton fluence and production of secondary particles that can have a relevant contribution to the total energy deposition. The contribution of nuclear reactions for proton beams in water is about 10% for 150 MeV and up to 20% for 250 MeV (Malmer 2001). Low-energy secondary particles are emitted isotropically in the centre-of-mass frame of the mother nucleus while the high-energy ones are originated mostly in the intra-nuclear cascade and pre-equilibrium stage and are emitted in the forward direction. Secondary emissions (e.g. de-excitation photons) are highly relevant to monitor the proton range in vivo, as investigated in this thesis.

2.2 Beam delivery and treatment planning in proton therapy

Protons can be accelerated by cyclotrons, synchrotrons or synchrocyclotrons, more details can be found in the work of (Owen et al. 2015).
There are two approaches to deliver dose to the target tumor using external beams, i.e. proton beams in this work.
In passive spreading techniques, the quasi mono-energetic proton beam provided by the

accelerator is spread laterally and longitudinally so that the dose is delivered homogeneously in the target. This is the so-called spread out Bragg peak (SOBP). Typically, a system consisting of scatterers with different complexity is used for lateral spreading, while the longitudinal range modulation is achieved by using either a stationary absorber (for small modulation), or a rotating absorber of an appropriately tailored profile (range modulator wheel) or plates of variable thickness [Schippers, 2015]. To deliver the dose in the assigned target volume, a patient specific combination of collimators and compensators is used (figure 2.4). This technique is limited by the neutron production due to the additional materials along the beam path and by the poor conformation in the proximal part of the target. The other one is a much more modern technique, named active scanning. The

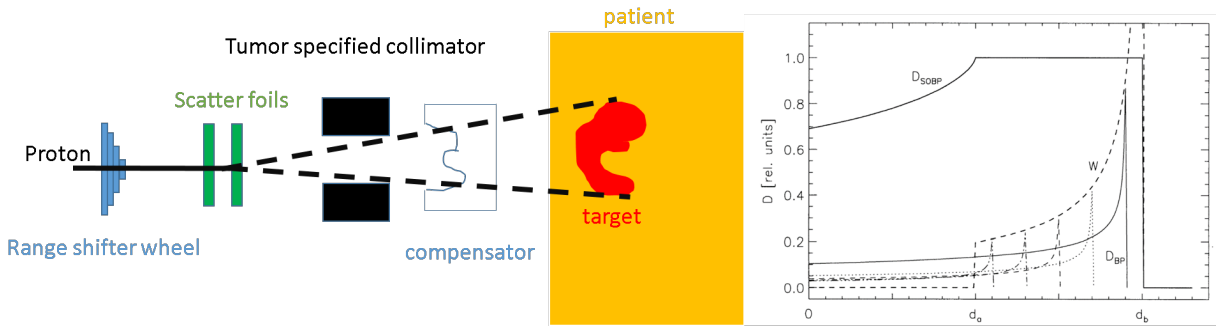


Figure 2.4: Left: Schematic setup of passive spreading techniques, in which the proton beam is modified using a tumour specific collimator to deliver the required dose in the target. Right: a typical SOBP can be achieved by a weighted superposition of elementary Bragg peaks in the desired range of depths (Bortfeld & Schlegel 1996).

proton beam is deflected and steered by dipole magnets to scan the target volume voxel by voxel (see figure 2.5). The penetration depth of the Bragg peak is adjusted by varying the energy of the beam (at the accelerator level in synchrotons or immediately after extraction in cyclotron-based systems). Compared to the passive spreading technique, this method shows advantages in terms of the proximal and distal dose conformity as well as reduced neutron production. Besides, this approach does not need a patient specific hardware and the protons are used in a more efficient way.
More recently, an advanced type of treatment planning with active scanning delivery, called intensity-modulated-particle-therapy (IMPT), has been well-developed and adopted worldwide (Rehman et al. 2018). In IMPT, various beam intensities resulting in highly heterogeneous dose distributions per field are generated intentionally in comparison to traditional ion beam therapy, where uniform dose distributions per field were delivered to the patient. This allows the IMPT technique to maximally exploit the degrees of freedom of the dose optimization process for either a final homogenous dose distribution (for the sum of the treatment fields), or to deliver varying radiation dose in selected target regions (Schlegel 2006).
To determine the energies, positions and intensities of the individual pencil beams (PB)

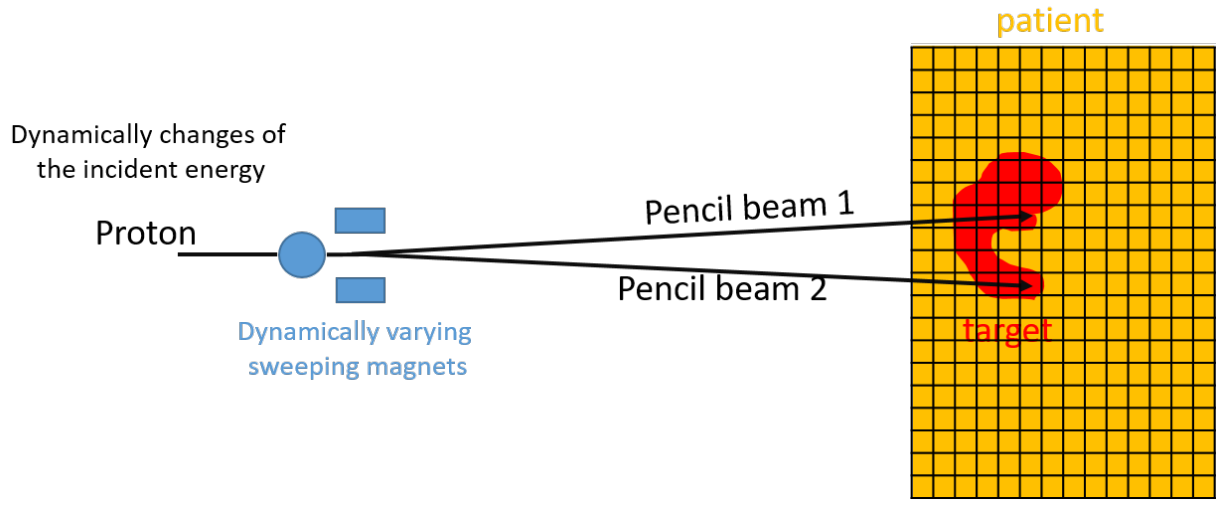


Figure 2.5: In the proton beam active scanning system, individual pencil beams are deflected by two orthogonal magnetic fields (dipoles) to scan all the voxels (3D grids) of the target. The intensity, direction and incident energy of each PB can be varied while no compensator, collimator, scatter foil or range shifter is needed.

used in IMPT, a treatment planning system (TPS) is needed to find the solution achieving the dose distribution closest to the set prescription. The main goal is to deliver the intended level of dose covering (typically homogenously) the target volume, while minimizing the dose delivered in healthy tissue and preserving special constraints on radiosensitive organs, so called organs at risk (OARs).
Generally, to make a treatment plan (TP), the patient X-ray computed tomography (CT) and the contours of clinically relevant structures are imported into the system. The CT numbers are in Hounsfield scale unit (HU):

$$HU = 1000\frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \tag{2.11}$$

where μ, μ_{air} and μ_{water} are the linear attenuation coefficients of the corresponding materials, air, and water, respectively. HU values for materials e.g. air, water and soft tissue are -1000, 0 and between 100-300, respectively (Feeman 2010).
In radiotherapy, the gross, palpable, visible or clinically demonstrable location and extent of the malignant growth is called gross tumour volume (GTV). The clinical target volume (CTV) includes a GTV and / or local subclinical spread of disease at a certain probability level that must be treated. The planning target volume (PTV) used for TP contains the CTV and an additional margin to account for the different types of variations and uncertainties (Jones 1994). The prescription dose distribution $D_{pre,j}$ (here j is the identifier of voxels) is set based on the clinical goal in the PTV, along with the dose constraints in the OARs. The dose distribution $D_{i,j}$ of all available individual PB_i is calculated using

a pencil beam algorithm. The treatment plan is then optimized by minimizing the cost function which describes the dose difference between the prescription and the treatment plan. An exemplary cost function in the TPS (CERR, cf. chapter 4) used in this thesis is:

$$c(w) = \sum_j p_j (w \cdot D_{i,j} - D_{j,pre})^2 \tag{2.12}$$

where w is the weight vector, i.e. the numbers of protons of all individual PBs in unit of 10^6, p_j is the penalty factor for voxel j. In this study p_j is set to 10000 for voxels in the PTV, while it is set to 10 for voxels in the OARs.
There are different software options for treatment planning algorithms. In this work, a particle extension of the research TPS CERR (A Computational Environment for Radiotherapy Research) (Deasy et al. 2003, Schell & Wilkens 2010, Resch et al. 2017) is used to create treatment plans. More details are presented in chapter 4.

2.3 The challenge of proton range uncertainties

One of the main advantages of proton therapy is the inverse energy deposition along the beam path enabling highly conformal dose delivery to the tumor and low toxicity rate in the healthy tissue (cf. chapter 1). On the other hand, the treatment accuracy is highly sensitive to potential differences between the actual dose delivery and the dose predicted by the TPS, especially due to uncertainties in the location of the Bragg peak within the patient, related to the proton beam range (see figure 2.6).

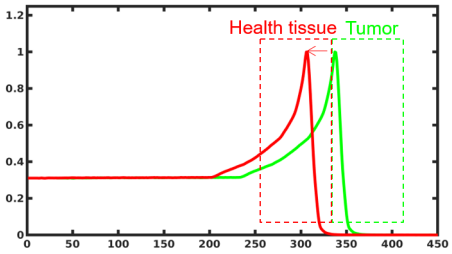


Figure 2.6: Schematic illustration of the effect of range uncertainty on the treatment delivery. The healthy tissue and the target, i.e. tumor, are represented by the red and green rectangle, respectively. In the TP, the Bragg peak of one PB is predicted to be delivered in the tumor (green profile). If this PB is affected by proton range uncertainty, i.e. the actual dose delivery changes to the red profile, most of the dose would be delivered in the healthy tissue instead of the target.

2.3.1 Sources of proton range uncertainties

In clinical proton beam therapy, non-negligible proton range uncertainties that can undermine the intended treatment accuracy are caused by many factors, which can be classified into two sources: treatment plan and treatment delivery.

Treatment plan

The TP is made using the X-ray CT scans of the patient, expressed in HU scale. Proton range uncertainties can be caused by not only the noise generated due to CT artefacts and beam hardening (Paganetti 2012), but also the semi-empirical conversion of the CT numbers into proton stopping power ratio in tissue relative to water (Schaffner & Pedroni 1998), leading to proton range inaccuracies of ca. 1-3% (Schaffner & Pedroni 1998), or even more. Analytical algorithms are typically used for dose calculation in conventional TPSs. The dose calculation mostly determines the proton beam range based on the water equivalent thickness of the patient, neglecting the position of heterogeneities relative to the Bragg peak (Paganetti 2012). The complexities of tissue geometries and densities thus affect the accuracy of the analytical dose calculation. This effect as a function of geometric complexity and proton energy was studied for example by (Sawakuchi et al. 2008), who reported an uncertainty of the falloffs of the dose profiles in the order of $\sim$2 mm for a 220 MeV proton beam. Nowadays, Monte Carlo (MC, cf. section 4.1) dose simulations are used within most advanced TPSs to overcome this effect. Using MC dose engines, a huge amount of incident particles (e.g. 10^5 particles per PB) is simulated in the heterogeneous tissue geometry and composition, passing through random path sampling and interaction from probability distributions calculated by the system, thereby typically achieving a more accurate dose calculation. A comparison between analytical and MC dose distributions in the same patient anatomy is shown in figure 2.7.

Treatment delivery

Usually delivery of the therapeutic treatment is fractionated over several weeks and thus performed at different time points, where the actual patient anatomy can be different from the planning CT (pCT). This may be due to not only potential anatomical changes caused by, e.g. weight loss, tumor shrinkage or tissue swelling, but also alignment issues due to e.g. positioning uncertainties in each treatment session. Additionally, patient anatomical locations can change within the treatment delivery due to e.g. breathing and heartbeats. Both inter- and intrafractional changes lead to unwanted proton range uncertainties.

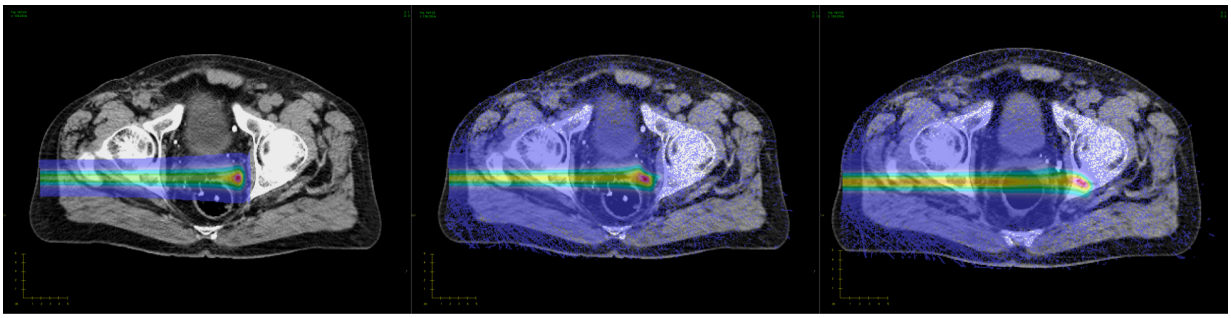


Figure 2.7: The dose calculated using a simple analytical dose engine of a research TPS CERR (Deasy et al. 2003, Resch et al. 2017) (cf. section 4.1) is shown on the left (note that the used CERR system does not implement pencil beam splitting, thus showing inferior performance than commercial TPSs), while the MC dose for the same PB is shown in the middle (on the planning CT, pCT). The right figure shows the MC dose calculation for the same PB in the case of a substantial anatomical change (on a different CT).

This effect is shown in figure 2.7 for an example of inter-fractional anatomical change.

Different approaches are applied to mitigate the planning uncertainties with robust treatment planning algorithm (Inaniwa et al. 2011) or improved image guidance to capture daily anatomy using e.g. cone beam CT (CBCT) (Venkatesh & Elluru 2017) or improvement of the stopping power estimation using e.g. dual energy CT (DECT) (Goo & Goo 2017) and proton CT (pCT)(Testa et al. 2013).
This work focuses on the in-vivo proton range verification using prompt gamma imaging.

2.3.2 Approaches for in-vivo range verification

To overcome the issue of proton range uncertainties, extensive research is ongoing to develop techniques that can enable an in-vivo monitoring of the actual proton beam range, ideally during delivery. This ambitious goal can be achieved by detecting the thermoacoustic signals accompanying proton irradiation of tissue (for pulsed beams only) or secondary particles, particularly energetic photons, generated in nuclear reactions, i.e. ionoacoustics, positron emission tomography (PET) and prompt gamma (PG) imaging. These techniques, at a different level of maturity and clinical translation, are briefly described in the following.

Ionoacoustics

Energy lost by pulsed ion beams penetrating a medium can cause a localized heating and a thermal expansion, generating detectable thermoacoustic emissions. By detecting the weak pressure waves (ionoacoustic signal), especially

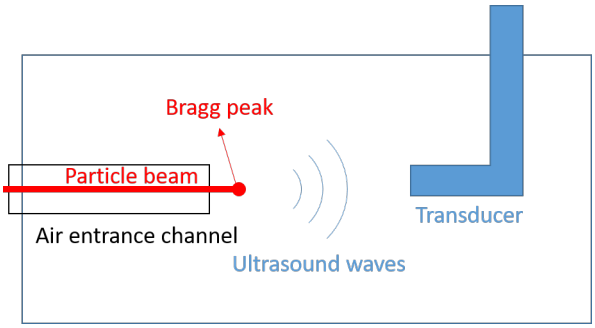


Figure 2.8: Schematic setup of ionoacoustics detection of pulsed proton beams in a water tank

using time-of-flight (TOF) measurements in combination with the knowledge of speed of sound in the traversed medium, the Bragg peak of a pulsed proton beam can be monitored in vivo (Parodi & Assmann 2015). A schematic setup of ionoacoustics is shown in figure 2.8. It was as early as 1979 that the feasibility of ionoacoustics was first investigated for beam diagnostics by (Sulak et al. 1979). In 1995, the first clinical investigation was conducted by (Hayakawa & Tsujii 1995) during the pulsed proton treatment of a liver cancer patient. However, due to the lack of tailored instrumentation and suboptimal passive proton delivery systems, the complexity of the ionoacoustic signals limited the performance of this proof-of-principle clinical investigation. The ionoacoustic technique reveals more potential nowadays because of the more favorable active beam delivery, along with improvements of ultrasound transducer technology as well as the commercial establishment of intrinsically pulsed proton accelerators (synchrocyclotrons).
The dose deposition profile can also be reconstructed by detecting and mathematically inverting the acoustic waves produced upon the energy loss of protons in water/tissue (Kellnberger et al. 2016). According to (Assmann et al. 2015), the Bragg peak position and the full width at half maximum (FWHM) of its Bragg peak width predicted by a MC simulation platform based on the code Geant4 (Geometry and Tracking 4, (Agostinelli et al. 2003)) were found in sub-millimeter agreement with 1D ionoacoustic measurements in water at a research 20 MeV pulsed proton beam. In the same homogeneous target (water), the accuracy of ionoacoustics-based proton range verification was found to be better better than 1.5 mm for pulsed delivery of 50 MeV protons, although in the case of high dose delivery (average of 1024 pulses, each delivering 2 Gy) for the considered experimental setup and detector technology (Patch et al. 2016). The feasibility of this technique was also proven more recently using a synchrocyclotron at clinical energies (Lehrack et al. 2017), where ionoacoustic range measurements in water showed sub-millimeter range accuracy and precision at a Bragg peak dose of about 10 Gy for 220 MeV initial proton beam energy.

Similar results were also reported for an artificially pulsed clinical cyclotron in (Jones et al. 2015). However, the power of this technique is still challenged by several issues, e.g. the requirement of dedicated instrumentation of high sensitivity and large bandwidth, the knowledge of the acoustic tissue properties and the difficult exploitation of complex signals caused by tissue heterogeneities. Nevertheless, the promising simulation studies in clinical scenarios that have been conducted recently still motivate the ongoing developments of this cost-effective technique for anatomical locations of easy acoustic access (e.g., prostate, breast), ideally in combination with diagnostic ultrasound imaging for co-registration of the Bragg peak to the underlying anatomy [Assmann et al., 2015, Hickling et al., 2018, Jones et al., 2018, Patch et al., 2018].

Positron emission tomography

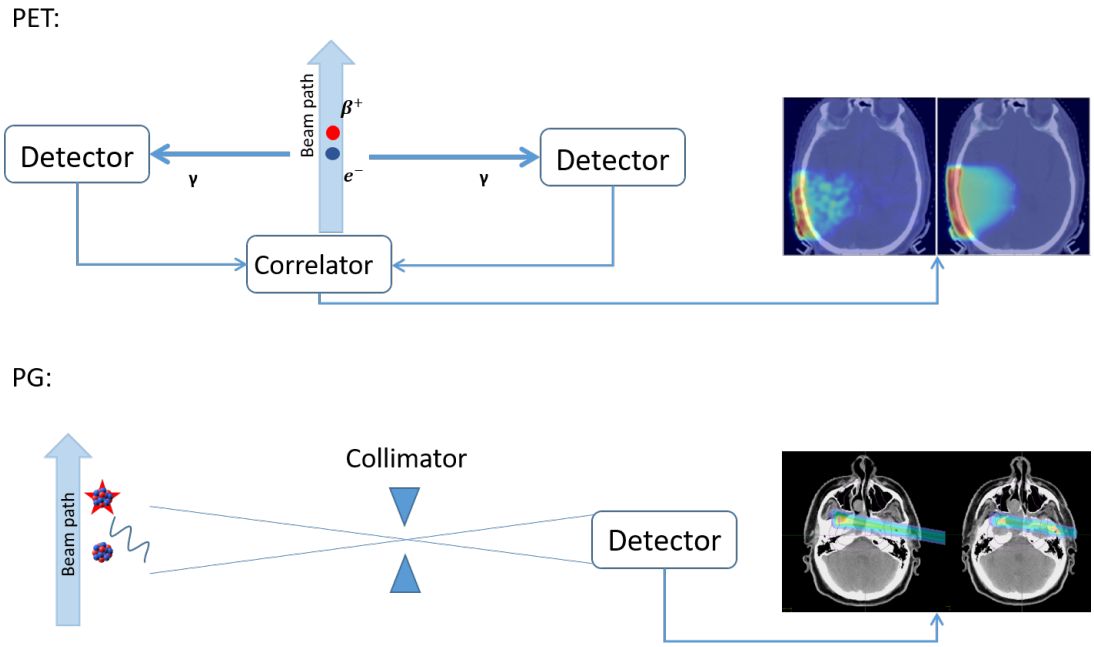


Figure 2.9: Schematic setups of PET and PG based proton range verification are shown on the top and bottom, respectively. Various types of PG systems are discussed in the following chapter. The exemplary PET imaging is from the work of (Remmele et al. 2011).

Using PET imaging for range verification of ion beam therapy was first proposed by (Maccabee et al. 1969). PET-based proton dose monitoring was then investigated by different groups during the 90s (Litzenberg et al. 1993, Vynckier et al. 1993, Paans & Schippers 1993, Oelfke et al. 1996, Litzenberg et al. 1999), and regained interest since the late 2000s.
In proton therapy, positron emitters are generated along the beam path through

nuclear fragmentation reactions, leaving a footprint of radiotracers that decay via positron emission (Zhu & Fakhri 2013). Two anti-parallel coincidence gammas of 511 keV are generated when the emitted positron travels a small distance and then annihilates with an electron in the medium. The falloff of the PET signal is well before the Bragg peak, around 4-6 mm (Parodi & Enghardt 2000). The rapid falloff of the dose delivery, i.e. proton range, can thus be inferred by the distribution of the positrons emission monitored by detecting those gammas using a PET detector (Parodi 2012). A schematic set up of a PET system is shown on the top of figure 2.9.
The spatial distribution and integral yield of several species of radionuclides (^{15}O, ^{11}C, ^{30}P, ^{38}K, etc) for clinically planned proton treatments for head and neck and paraspinal tumors was calculated using MC simulation by (Parodi et al. 2007a, Parodi et al. 2007b). The most important radionuclides produced in the case of soft tissues are ^{11}C, ^{13}N and ^{15}O. The ^{15}O is the dominant radionuclide produced during irradiation due to the high oxygen density, but rapidly decays due to the half-life of only $\sim$2 min, while ^{11}C contributes more to the activity measured after irradiation, because of its longer half live of 20.3 mins (Valverde et al. 2018). Hence, the PET signal depends on the timing of the data acquisition and underlying contributing isotopes during or after irradiation, when using PET scanners located inside or outside the treatment room. However, due to the different physical processes involved, the dose distribution cannot be directly correlated with the positron activity. Moreover, depending on the time elapsed between irradiation and imaging, the PET-dose correlation can be additionally challenged by the low signal-to-noise ratio and distortions caused by physiological processes, so called washout.
To perform the PET-based proton range verification, the PET measurement estimation is firstly created by either using analytical methods (Parodi & Bortfeld 2006, Miyatake et al. 2011) or MC simulation, which is the most accurate method and was firstly proposed and applied on 9 patients in a clinical study using a remote PET/CT scanner in (Parodi et al. 2007b). To this end, the simulated proton fluence spectrum, according to the detailed MC simulations of the TP, was convolved with the energy dependent cross section for the main reaction channels and then scaled with the target nuclei densities deduced from the patient CT. Corrections for biological washout and radioactive decay were applied taking into account the time of irradiation and offline data acquisition. The simulated PET image was finally obtained using a convolution kernel to model the resolution of the scanner, and then compared to the measured image, reconstructed by the commercial scanner, to estimate whether the dose was delivered as planned. Range verification accuracy of 1-2 mm was achieved in head and neck cancer patients in the case of good immobilization and registration between treatment and imaging CT as well as reduced washout in bony structures (Parodi et al. 2007b). In the cases of other anatomical sites, an accuracy of only 3-5 mm was found, indicating a lack of millimeter range

accuracy for every tumour indication (Knopf & Lomax 2013). Recently, a dedicated in-beam dual head PET scanner has been developed to overcome the limitations of offline PET imaging and to detect the activity during irradiation (Ferrero et al. 2018), with a first clinical evaluation ongoing.

Prompt gamma imaging

Nuclear excitations and related PG emissions due to inelastic interactions of incident protons and nuclei in the beam path occur until the energy of primary protons decrease below the reaction threshold, which is typically a few millimeters before the Bragg peak, resulting in a falloff close to the Bragg peak in the case of PG(Knopf & Lomax 2013) (cf. figure 2.10). This is similar to what described for PET, but the energy threshold is typically lower. The energetic PGs are emitted by the de-excitation of the excited nuclei to the lower energy states (cf. table 2.1). Discrete gamma emissions can be extracted from the standard measurements (Verburg et al. 2012) when the cross sections of the transitions of the quantized energy states of excited nuclei are high enough. This process strongly depends on the energy of the incident protons and the tissue composition of the target. The gamma lines resulting from the interaction between protons and the most abundant elements in human tissue, i.e. ^{12}C, ^{16}O, and ^{14}N, are shown in table 2.1 (Kozlovsky et al. 2002, Verburg et al. 2012). A schematic setup of a PG system is shown on the bottom of figure 2.9, and more details of proposed detection schemes will be given in chapter 3.

For ^{12}C, the 4.44 MeV line is the dominant contribution. For levels above the 4.44 MeV line, most of the ^{12}C decays through alpha emission (Ajzenberg-Selove 1990). For ^{16}O, most of the gamma emissions are contributed by the lines at 6.13, 6.92 and 7.12 MeV in the case of low energy protons (Tilley et al. 1993), while the contribution of the ^{16}O(p,x)^{12}C* reaction channel increases due to the 4.44 MeV PGs from the residual ^{12}C*. Due to the typically lower abundance in tissues, compared to ^{12}C and ^{16}O, ^{14}N contributes less to PG emission.
Compared to PET based proton range verification, PG has higher production rates (especially in the case of a therapeutic dose rate of 2Gy/min) (Moteabbed et al. 2011). Moreover, PG needs shorter time for detection (within a few nanoseconds after the nuclear interaction induced by the irradiation), hence allowing good on-line monitoring if detectors of sufficient detection efficiency are available. The PG signal is also better correlated with the primary range because the cross sections for PG production have typically lower energy thresholds than for positron emitter yields (Moteabbed et al. 2011). Finally, the positron emitted in the β^{+}-decay has a certain range in tissue before annihilating with an electron, thus intrinsically limiting the achievable spatial resolution

Target elements	Reaction channel	E_γ (MeV)	Transition	Mean life (s)
^{12}C	$^{12}C(p,x)^{11}C^*$	2.00	$\frac{1}{2}^-$ 2.00 $\rightarrow$ $\frac{3}{2}^-$ g.s.	1.0×10^{-14}
	$^{12}C(p,p')^{12}C^*$	4.44	$2^+ \rightarrow$ 4.44 0^+ g.s.	6.1×10^{-14}
^{16}O	$^{16}O(p,p')^{16}C^*$	2.74	2^- 8.87 $\rightarrow$ 3^- 6.13	1.8×10^{-13}
		6.13	3^- 6.13 $\rightarrow$ 0^+ g.s.	2.7×10^{-11}
		6.92	2^+ 6.92 $\rightarrow$ 0^+ g.s.	6.8×10^{-14}
		7.12	1^- 7.12 $\rightarrow$ 0^+ g.s.	1.2×10^{-19}
	$^{16}O(p,x)^{12}C^*$	4.44	$2^+ \rightarrow$ 4.44 0^+ g.s.	6.1×10^{-14}
	$^{16}O(p,x)^{15}N^*$	5.27	$\frac{5}{2}^+$ 5.27 $\rightarrow$ $\frac{1}{2}^-$ g.s.	2.6×10^{-12}
^{14}N	$^{14}N(p,p')^{14}N^*$	1.64	1^+ 3.95 $\rightarrow$ 0^+ 2.31	6.9×10^{-15}
		2.31	0^+ 2.31 $\rightarrow$ 1^+ g.s.	6.9×10^{-15}
		5.11	2^- 5.11 $\rightarrow$ 1^+ g.s.	6.3×10^{-12}

Table 2.1: Reaction channels of relevant PG emissions in proton therapy (g.s. ground state) (Kozlovsky et al. 2002, Verburg et al. 2012).

of PET imaging.

Collimated detection of PG was firstly explored in an experiment by (Min et al. 2006), in which the distance between the Bragg peak and the maximum of the PG distribution was within 1-2 mm in the case of protons with 100 MeV initial energy stopped in water. Due to the deeper penetration depth and higher neutron background, this distance increased in the cases of higher energy protons (150 and 200 MeV). Different studies, both experimental and simulation, have been conducted to further explore the feasibility of PG-based proton range verification in the successive years (Polf et al. 2009a, Polf et al. 2009b, Kang & Kim 2009, Frandes et al. 2010, Richard et al. 2011, Draeger et al. 2018). The first clinical PG imaging was conducted in a study of passively scattered proton therapy of a head and neck cancer patient at OncoRay in 2016, Dresden, Germany and a global inter-fractional range shift in the order of 2.0 mm was detected (Richter et al. 2016). In 2017, the first clinical spot-by-spot PG imaging for active scanning was conducted for a relatively homogeneous case of a brain tumor by (Xie et al. 2017), demonstrating that a retrieval precision of 2 mm can be achieved. Since then, this is a very active field of research, as reviewed in the next chapter.

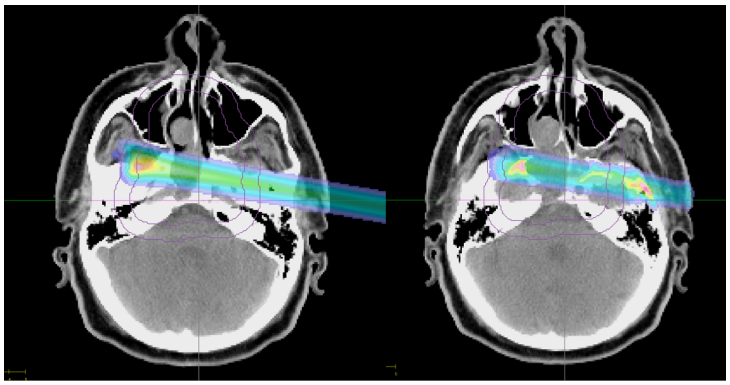


Figure 2.10: The distributions of the MC calculated dose and corresponding PG emission are shown on the left and right, respectively.

Chapter 3

PG range monitoring

3.1 PG-based proton range verification systems

PG-based proton range verification cannot rely on standard nuclear medicine instrumentation tailored to low-energy (typically 140 keV) single photon detection, but requires dedicated cameras to overcome the challenges of a broad- and high-energy distribution in the multi-MeV range, high instantaneous count rate and compatibility constraints with patient irradiation. Several types of PG devices have been proposed over the last decade to retrieve directional information of the PG emission, correlated with the beam range, exploiting either mechanical or electronic collimation (Krimmer et al. 2018). Regardless of the chosen detection scheme, a considerable background resulting from the simultaneously created neutrons may limit the signal-to-noise ratio of PG imaging. This is caused by direct interactions of the neutrons in the detectors and/or from photons produced by the neutrons within the collimators and/or other structures including the patient surrounding the detectors. Neutrons directional and spatial correlation with the proton range is typically lost due to neutron scattering in the patient and surrounding. Thus, the effect of the neutrons on the PG signal can only be detrimental and needs to be suppressed. To this end, (Biegun et al. 2012) proposed to use time of flight (TOF) detection to reject neutron background. Their MC simulation showed that using a shifting time window with a width of TOF of 1.0 ns can significantly reduce the neutron background (more than 99%), especially in the case of proton beams with a microstructure that is typical of cyclotron accelerators. In addition to the timing properties of the PG, which can also be exploited to retrieve proton range information or as a global monitor of the applied treatment, also their energy is a characteristic feature specific to the involved nuclear de-excitation process. The idea to exploit the energy information of PG was firstly proposed and investigated by (Polf et al. 2009a). In their pioneering simulation and experimental studies, it was shown that it is possible to distinguish the characteristic PG emission lines from the major elemental constituent atoms (C, O, Ca) during proton dose delivery. Hence, recent systems aimed to combine directional detection with spectroscopic PG information to increase the robustness of the proton range verification approach, additionally offering the possibility

to recover the main carbon and oxygen tissue composition in the region of interest imaged by the system. A summary of all these detection technologies is given in the following.

3.1.1 PG detection prototypes

Collimated systems

Pinhole/knife edge cameras

The first feasibility study of PG range verification has been performed by (Min et al. 2006), using a single CsI(Tl) scintillation detector behind a collimator with a parallel slit geometry. Measurements have been performed irradiating a water target with mono-energetic proton beams in the 100-200 MeV energy range, and a dedicated collimator (at 90° from beam direction) has been designed to moderate and capture fast neutrons as well as to prevent unwanted prompt gamma rays not perpendicular to be the beam direction to reach the scintillation detector. The PG profile has been obtained via a scan of the target in front of the collimator detector arrangement (sensitive to the collimator alignment to the detector). A promising correlation of the distal falloff region of the PG distribution with the additionally measured depth dose distribution has been observed (Min et al. 2006).
Another promising experiment using a CsI(Tl) scintillation detector behind a pinhole aperture for high energy PG detection was performed by (Kim 2009). The idea is to use the same principle as an optical pinhole camera for high energy PG detection. Similar concepts of collimated systems were then investigated and optimized in extensive Monte Carlo simulation studies by different groups, e.g. (Bom et al. 2011). It was in 2012 that the first prototype of a knife-edge camera, consisting of one lutetium-yttrium oxyorthosilicate (LYSO) crystal behind a slit collimator (Peloso et al. 2010, Busca et al. 2010), was built by the company IBA in collaboration with Politecnico di Milano (Smeets et al. 2012). This design is based on the HiCam system, a compact Anger gamma camera developed in the framework of the HiCam (high resolution camera) project (Fiorini et al. 2008, Peloso et al. 2010), which can achieve a very good intrinsic resolution of 0.8 mm FWHM when imaging 140 keV gamma rays from ^{99}Tc (Smeets et al. 2012). The initial HiCam system was designed for nuclear medicine applications where both high resolution and compactness are required, such as diagnosis of brain tumours, breast and thyroid cancer as well as small rodents imaging for the evaluation of pharmaceutical distribution or functional effects resulting from medical treatments (Busca et al. 2010). The second generation of this prototype was established by using LYSO scintillator slabs read out by silicon photomultipliers in order to improve the system in terms of counting statistics and photon-detection efficiency at the higher energies of interest for PG monitoring. The retrieved PG measurement is sensitive to the position of the camera, which means that the alignment should be performed with respect to the patient along the beam direction instead of the beam isocenter.
Although originally designed for the PB scanning delivery technique (see chapter 2), the first clinical implementation of the IBA knife-edge camera was conducted in 2016 for passively scattered proton therapy of a head and neck cancer patient at OncoRay in Dresden, Germany. A global inter-fractional range shift in the order of 2.0 mm was detected in the

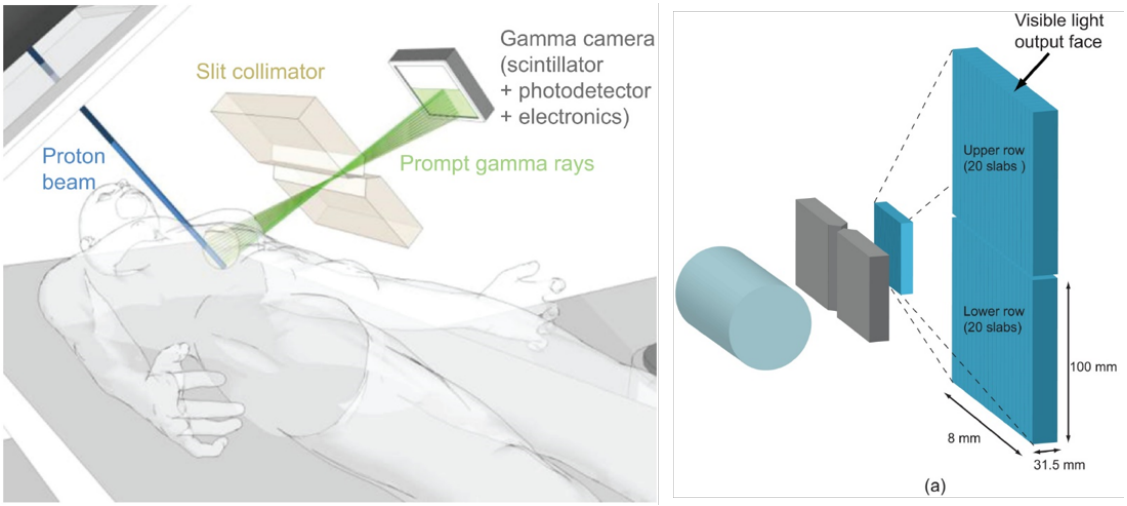


Figure 3.1: The left panel shows the concept of the slit camera: a slit collimation gives a 1D projection of PG emission along the beam path on a scintillation detector. The configuration of this system is shown on the right panel (Perali et al. 2014).

monitored 7 fractions, and was found in good agreement with the information deduced from dose recalculations on in-room control CTs (Richter et al. 2016). The second clinical implementation was performed by (Xie et al. 2017) in PB scanning proton therapy for a brain cancer patient at the University of Pennsylvania. The detected average range shift was within 1.0-2.0 mm, which is below the fixed distal margin of 5.0 mm applied clinically. The retrieval precision was reported to be affected by the amount of delivered protons, and a proton statistics (i.e. number of protons) of 2×10^8 was postulated to be needed for a reliable monitoring, which could be achieved by aggregating the collected signal from nearby PBs in the conventional TP (referred to as spot aggregation, explained more in detail in chapter 5).
A schematic setup is shown in figure 3.1. In this thesis work, the detection of PG by a slit camera is simulated using the Matlab-based toolkit REGistration Graphical User Interface (REGGUI) developed by the company IBA (Janssens et al. 2011). Relevant details will be given in chapter 4.

Multi-slit camera

This approach is an evolution of the single slit camera, i.e. using multiple detectors behind parallel slit collimators placed perpendicular to the proton beam direction (schematic setup shown in figure 3.2). PGs emitted from the target would pass through the corresponding collimator slit (sensitive to the position of the detector). The detectors are typically able to resolve energy and thus the resulting counts in a selected energy window (e.g. 1.1-10 MeV in the work of (Krimmer et al. 2015)) in the detector are registered. The measured PG distribution can be correlated to the distal falloff of the dose and with the knowledge of the 2D position of the beam entrance in pencil beam scanning delivery (from the TP or a fast detector in the beam), the 1D longitudinal PG information is complemented (Min

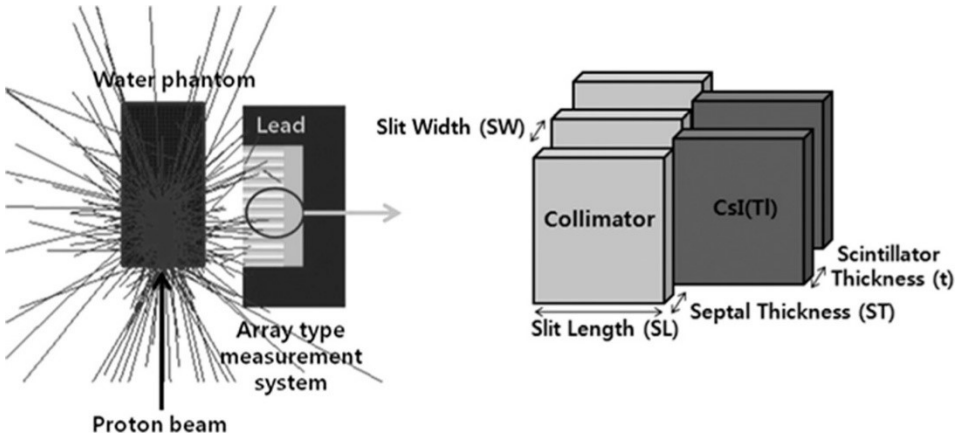


Figure 3.2: Schematic setup of a multi-slit camera system for PG measurement (Min et al. 2012).

et al. 2012, Krimmer et al. 2015). Measurements of PG generation as a function of proton penetration depth in the target have been conducted using crystal scintillators by (Pinto et al. 2015) and using semiconductor detectors by (Kelleter et al. 2017). By using these multi-slit camera systems, the field of view can be enlarged with respect to the previous single-slit solutions to encompass the complete beam path. Compared to other detection systems described in the next section, the multi-slit camera has good spatial resolution and does not rely on a reconstruction algorithm, although the signals could be blurred by inter-detector scattering (Pinto et al. 2014). According to the work of (Park et al. 2019), the signal quality of this camera prototype can be further improved by (1) using more detector modules, (2) increasing the size of the scintillator, (3) decreasing the length of the slit of the multi slit collimator.

Compton camera

A Compton camera (CC) system consists of one or multiple scatterer detectors, typically thin enough to enable only one scattering event per layer, usually followed by a thicker absorber detector of high atomic number. Incident gammas thus primarily undergo Compton interaction in the scatterer detector and are ideally fully absorbed with the photoelectric effect in the absorber detector. The kinematics of Compton scattering is used by this system to construct a source imaging without the use of collimators or masks (Phillips 1995). The direction of the incident gamma is restricted to the surface of a cone with an aperture equal to the Compton scattering angle. This information is used for the 3D reconstruction of the PG emission.

Compton telescopes were first built in the 1970s for astronomical observations. In 1974 the first CC for medical imaging was proposed by (Todd et al. 1974) and then investigated by (Everett et al. 1977) using a simplified detector prototype with two segmented semiconductor layers. In the last decade, Compton cameras have attracted much attention for in vivo proton range verification. The idea is to exploit Compton kinematics to restrict the direction of the incident photon to a cone based on the energy deposition and interaction

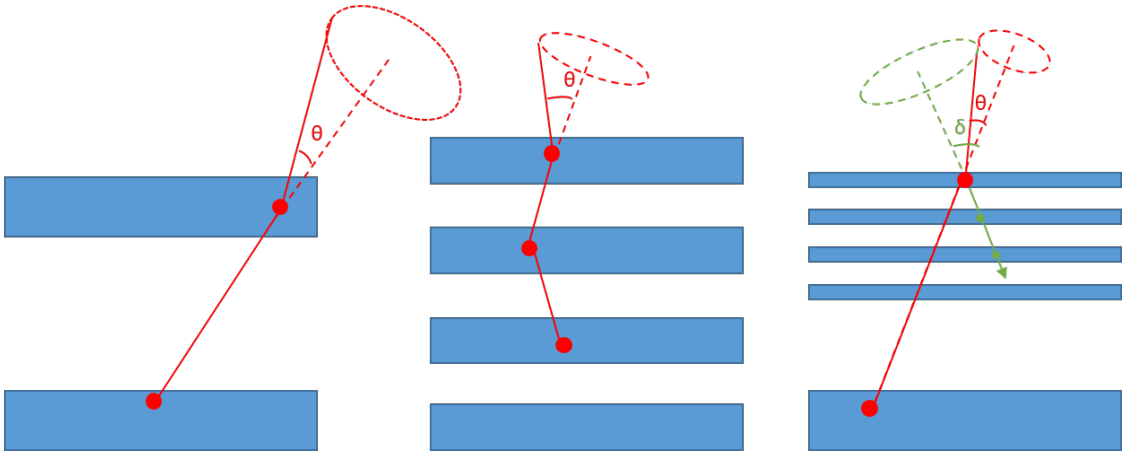


Figure 3.3: Schematic configurations of CC detectors. The left and middle panel shows a classic CC consisting of two and multiple detectors, respectively. On the right is the electron tracking configuration consisting of several thin layers able to register the path of the recoil electron. Ideally, the Compton scattered gamma interacts in a thicker detector component for full absorption, and the initial direction may be completely determined (δ and θ represent the total scattering angle and the photon scattering angle, respectively).

points in segmented or position sensitive monolithic detectors. Thus, the vertex of the photon generation is obtained via the superposition of multiple cones. This allows Compton cameras to create 3D PG information. Besides, in the relevant PG energy range of several MeV, Compton scattering is the dominant process and Compton cameras are well adapted in this case for the purpose of prompt gamma-ray detection (Krimmer et al. 2018).
The simplest Compton camera uses two detectors as shown in figure 3.3 left. This kind of Compton camera requires a total absorption of the scattered photon in the absorber for correct kinematics reconstruction, since the energy of the produced PG is a priori not known. However, this full absorption cannot always be guaranteed at the typical high energies of PG produced in proton therapy. Algorithms have been developed to use the elemental composition of the target to constrain the energy of the PG to a few discrete lines (Xu & He 2007, Draeger et al. 2016). An alternative (figure 3.3 middle) is to use two (or more) scatterers, so the kinematics of the incident PG can be completely determined and full absorption is no longer needed, although the efficiency is reduced by at least an order of magnitude (Roellinghoff et al. 2014). In contrast, the usage of multiple thin scatterer detectors as shown in the right panel of figure 3.3 can additionally enable the measurement of the recoil electron direction. This provides information about the total scatter angle that restricts the origin of the incident PG to an arc segment, again not requiring full absorption. Despite the good correlation shown for the measured PGs with the proton range, this approach is limited by the low efficiency of multiple interactions (Kurosawa et al. 2012).
Different CC prototypes have been developed using different detector technologies based on scintillators or solid state materials, e.g. the Polaris-J CC developed by H3D Inc.

for the group of Prof. Polf at University of Maryland (McCleskey et al. 2015, Polf & Parodi 2015, Draeger et al. 2018) and the MACACO CC developed by (Llosa et al. 2016) at instituto de Fisica Corpuscular, Valencia, Spain and CERN, Meyrin, Switzerland. This promising field is still in progress to allow the full deployment of CC into the clinic by e.g. improving count rate (Draeger et al. 2018) and reducing the cost (Krimmer et al. 2018).

Integrated yields counting systems

Prompt gamma timing (PGT)

PGs are emitted throughout the penetration process of protons in matter, as long as the energy is above the threshold for nuclear interaction. The proton transit time depends on the proton range (1-2ns for a range of 5-20 cm in the patient's body) and is reflected in the width of PG TOF distribution. The idea of monitoring proton range by measuring only the uncollimated PG TOF with respect to a given start signal (corresponding to the time the protons enter the target) was first proposed and investigated by (Golnik et al. 2014) using 150 MeV protons impinging on a graphite target. A fast scintillator detector, GAGG:CE (cerium-doped Gd3Al2Ga3O12) scintillator (Kamada et al. 2012) coupled onto a Photonics XP2972 photo multiplier tube (PMT), was used in this work as primary detector. The feasibility of this method was investigated in terms of the effect of changes in e.g. the target position, target thickness, incident proton energy and target components (air and bone) on the TOF distribution. The schematic configuration of the PGT method is shown in figure 3.4. The setup of the mentioned study is shown in figure 3.5.

In the case of high statistics, i.e. 10^{10} primary protons, range differences of 2.0 mm were

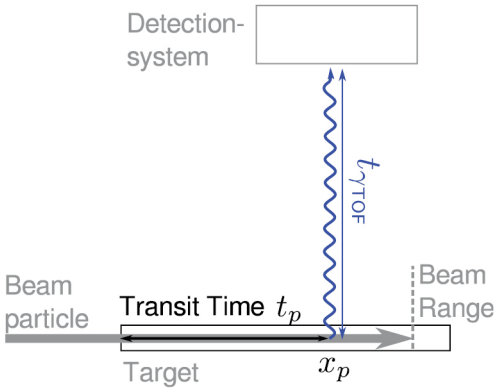


Figure 3.4: Schematic configuration of PGT. A PG is emitted along the beam track at x_p. The time t_p has elapsed since the corresponding particle entered the target. The emission of the PG is assumed to be instantaneous. The $t_{\gamma TOF}$ is the time for the PG reach the detector, which is largely dominated by the transit time. A detection system is measuring the time difference between the time of the particle entering the target and the arrival time of the PG at the detector. The measured distribution is denoted as PGT spectrum (Golnik et al. 2014).

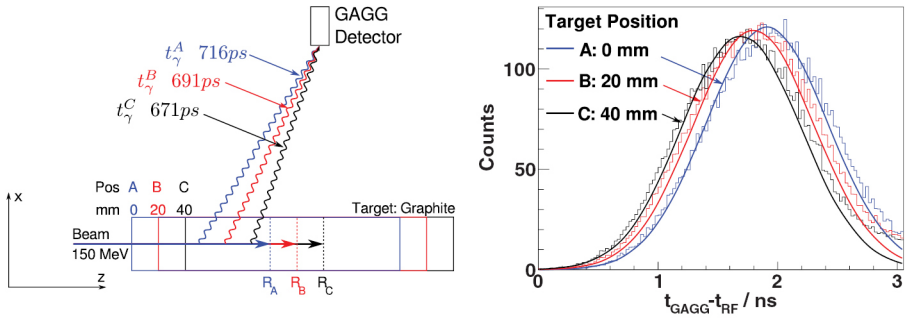


Figure 3.5: 150 MeV proton beams are delivered in a graphite target (density: 1.7 g/cm^3) at three different positions A, B and C, as shown on the left. The corresponding measured PG timing distributions, detected in this case by a GAGG detector, are shown on the right (Golnik et al. 2014).

resolved in homogeneous phantoms. The minimum range shift detected is less than 5.0 mm in the case of reduced statistics (Hueso-Gonzalez et al. 2015). The possible clinical implementation of the PGT method has been investigated at the University Proton Therapy Dresden using a pencil beam scanning plan (Werner et al. 2019). The proton range changes caused by material variation could be identified from the corresponding PGT spectra. An assignment of the PGT data to the individual PB allowed a spot-wise analysis of the variation of the PGT distribution mean and width, corresponding to range shifts produced by different air cavities (Werner et al. 2019). However, an excellent time resolution less than 1 ns is needed. This value might already be degraded by the time structure of the beam, especially with the accelerators of synchrocyclotron type (Krimmer et al. 2017).

Prompt gamma peak integral (PGPI)

To detect gross deviations in treatment delivery of proton therapy, (Krimmer et al. 2017) investigated the peak integrals of the distribution of uncollimated PG TOF at Centre Antonie Lacassagne in Nice, France. This approach enables to use an affordable and independent monitoring system to retrieve the TOF information, providing the possibility to discriminate the PGs generated in the beam line from those originated in the target. The combination of signals from multiple detectors placed around a target can be used to detect a misplacement of the target or provide a precise number of the registered PG that is independent of the target position (Krimmer et al. 2017).

Experiments showed that the PG peak integrals (PGPI) are sensitive to variations in the registered PG ray counts. A range shift of 3.0 mm in PMMA in the case of 10^8 protons can be detected by using the detectors covering a solid angle of 25 msr (Krimmer et al. 2017). A schematic setup is shown in figure 3.6.

Prompt gamma spectroscopy (PGS)

The prompt gamma spectroscopy (PGS) technique verifies the proton range by matching the energy- and time resolved PG measurements with models based on experimentally determined reaction cross sections (Verburg et al. 2013). The discrete gamma lines, attributed to specific nuclear level transitions, are identified in this approach, providing several benefits to improve the accuracy and sensitivity of range verification with a collimated PG

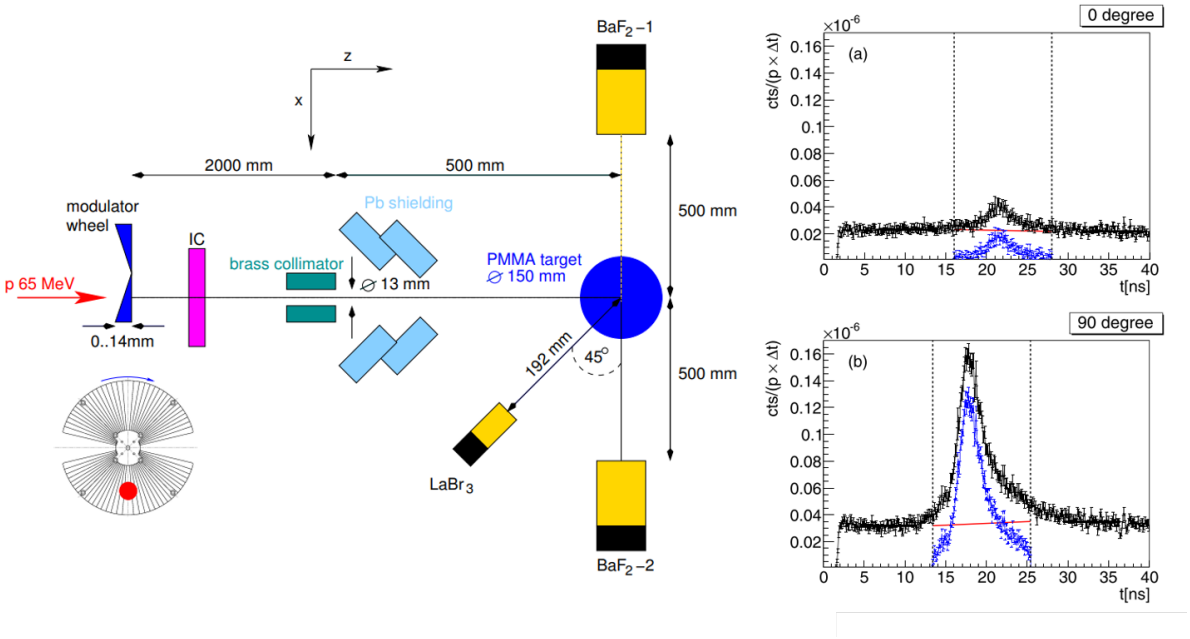
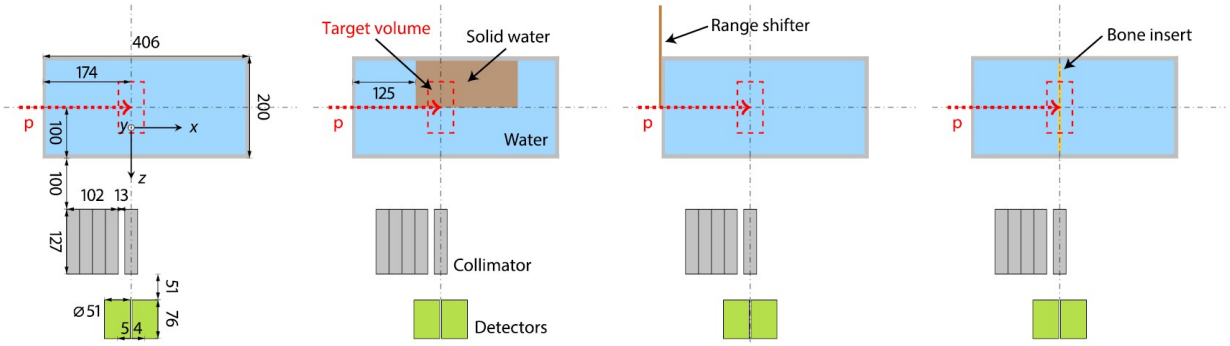


Figure 3.6: Configuration for the measurements of PGPI with cerium doped $LaBr_3$ and BaF_2 scintillation detectors (left) and the TOF spectra corresponding to different positions (degrees of rotation) of the modulator wheel (right). After subtracting the background (red line), the signals (blue) corresponding to the PG emission from the target are integrated within the limits (vertical dashed lines) (Krimmer et al. 2017).

detection system. First, each of these discrete PG emissions has a unique correlation to the proton energy and the cross sections of nuclear reaction that can be used in the proton range verification. Second, it provides the possibility to estimate the elemental concentrations of the target, making the range verification more robust if the proton beam stops in a tissue with an uncertain composition (Verburg et al. 2013).

The feasibility of this method has been demonstrated by (Verburg 2015) for passively scattered proton beams and by (Hueso-Gonzalez et al. 2018) for pencil beam scanning. In particular, the results of the latter experiment (setup shown in figure 3.7) demonstrate for the first time that a proton range verification precision of 1 mm is achieved in phantoms under clinically realistic conditions. The detection scheme is still mechanical collimated but with additional good energy resolution (provided by $LaBr_3$) of the detector and TOF for background suppression Hence, a new full-scale clinical prototype detection system based on PG spectroscopy has been developed and will be deployed for the clinical study with patients soon, see figure 3.8.



(a) Nominal **(b)** Solid water insert **(c)** 2.2; 5.2 mm shifter **(d)** Inner; Cortical bone

Figure 3.7: Schematic of the experimental setup of the PGS system demonstrated in (Hueso-Gonzalez et al. 2018). Protons (red arrow) irradiated a water phantom with different settings (a. no range error; b. solid water block covering half of the field; c. 2.2 and 5.2 mm water equivalent range shifters covering half of the field; d. 5mm thick slab of inner or SB3 cortical bone inserted in the middle of the field). The PG are collimated (gray) and measured with scintillation detectors of excellent energy and timing resolution (green).

3.2 Dependence of PG signal with counting statistics

The precision of the retrieved range from the PG signal is sensitive to the Poisson noise in the detected profiles. Spot-by-spot PG simulation using the IBA knife-edge system based on LYSO scintillating crystals conducted by (Xie et al. 2017) pointed out that a shift retrieval precision of 2 mm can be achieved when the number of protons is more than 2×10^8 (figure 3.9). The range resolving ability of a CC based on 3 layers of semiconductor detectors was investigated by (Draeger et al. 2018) during the irradiation of a tissue-equivalent plastic phantom with proton PBs for clinical doses delivered at clinical dose rates. The result showed that 3D images of PG emission can be produced with the delivery of as few as 1×10^8 protons and shifts in the proton beam ranges as small as 3 mm can be detected. As expected, profiles generated using 1e8 protons are much noisier than those generated using a higher statistics of 6.29×10^8 protons (Draeger et al. 2018) (figure 3.10 and 3.11).

In this thesis work, Geant4 simulations are employed to explore the correlation between the statistics of incident protons and the precision of PG collected by a phase space detector surrounding the target, which records all PG hit events in terms of position, energy, momentum and TOF. Relevant details of the code and results will be shown in section 4.1 and 5.1, respectively.

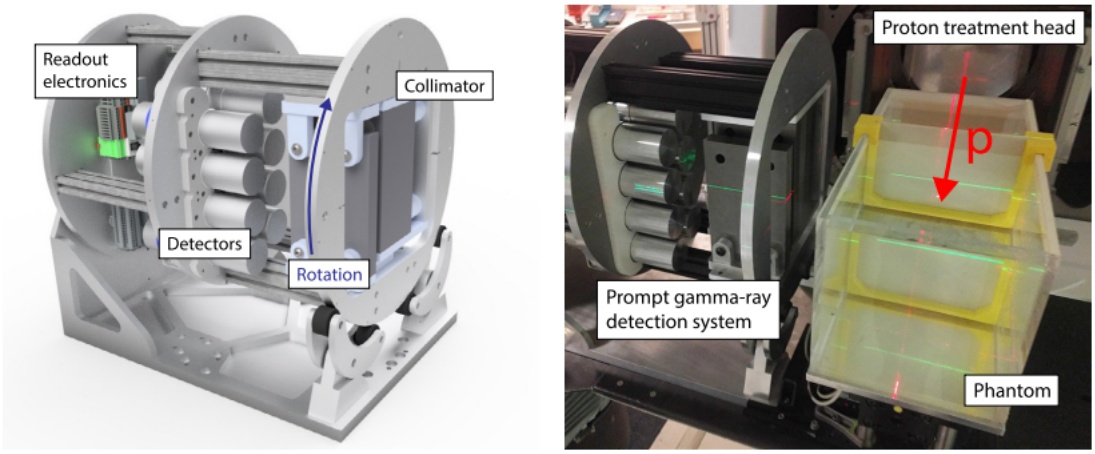


Figure 3.8: Left: 3D model of the clinical PGS prototype system. Right: photo of the system in the gantry treatment room. The red arrow shows the proton beam incidence direction. (Hueso-Gonzalez et al. 2018)

3.3 Influence of tissue heterogeneities on PG-based proton range verification

In addition to counting statistics, the accuracy of the PG based proton range verification can be affected by tissue heterogeneities, which may significantly alter the shape of PG distributions and change the correlation between the proton range and the PG depth profile.
This was experimentally investigated by (Priegnitz et al. 2015) using the IBA knife-edge slit camera (see section 3.2.1) during proton irradiation for tissue-equivalent targets. The PG range monitoring fails when the PG distal falloff is placed near pronounced density gradients, i.e. the beam stops in close vicinity to air cavities or low density lung tissue (see figure 3.12) (Priegnitz et al. 2015).
Similar effects were also reported in the MC study of (Schmid et al. 2015). The Bragg peaks may be distorted when there are large transversal tissue heterogeneities close to the end of the proton range and the shift calculation for PG profiles might be unreliable due the the changes of the falloff pattern caused by the complexity of the inhomogeneous tissue (see figure 3.13 and 3.14). In this case, the correlation between the falloffs of PG and dose profiles would be poor.

In this thesis work, the correlation between the falloffs of dose and PG as well as the effect of tissue heterogeneities on the dose delivery are investigated and quantified, as explicitly addressed in chapter 5 and 6.

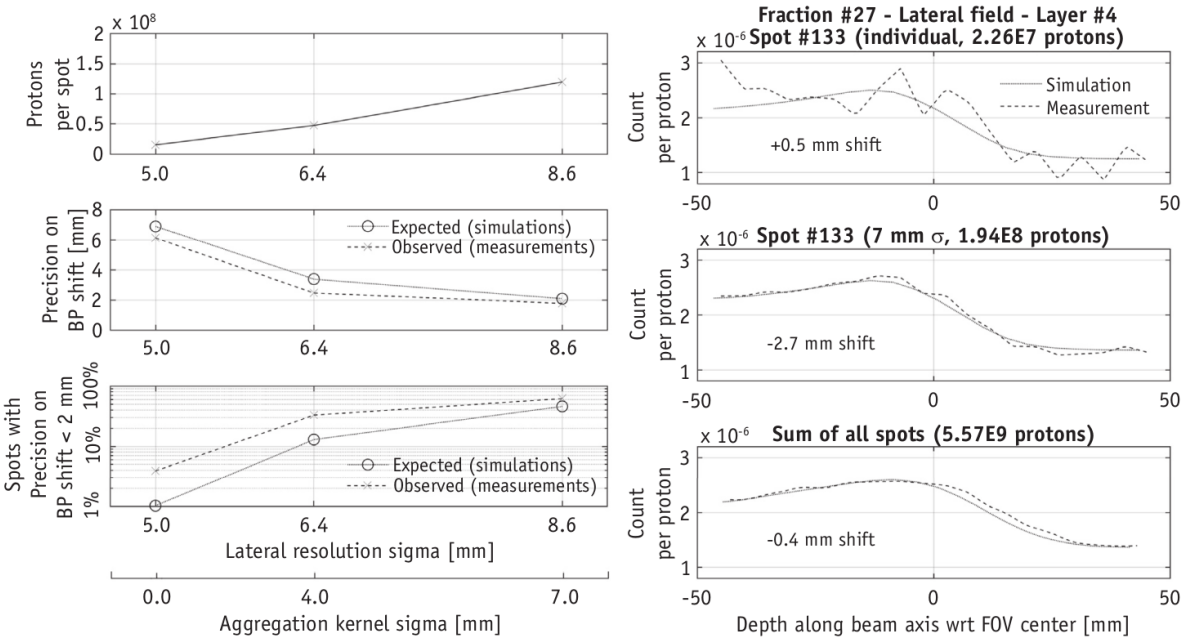


Figure 3.9: (Left) Precision of BP shift retrieval according to the evaluation presented in (Xie et al. 2017) for the IBA knife-edge camera. Top: number of protons per PB spot. Middle: expected and observed shift retrieval precision. Bottom: fraction of spots achieving a shift retrieval precision better than 2 mm as a function of lateral spatial resolution, defined as the sum in quadrature of the beam sigma and the sigma of the aggregation kernel explained in section 5.2. (Right) Example of the comparison between the measured and simulated PG profiles for an individual PB spot (top), for the aggregated spots (with an aggregation kernel of 7 mm sigma, middle) and for the entire energy layer including all PB spots (bottom) (Xie et al. 2017).

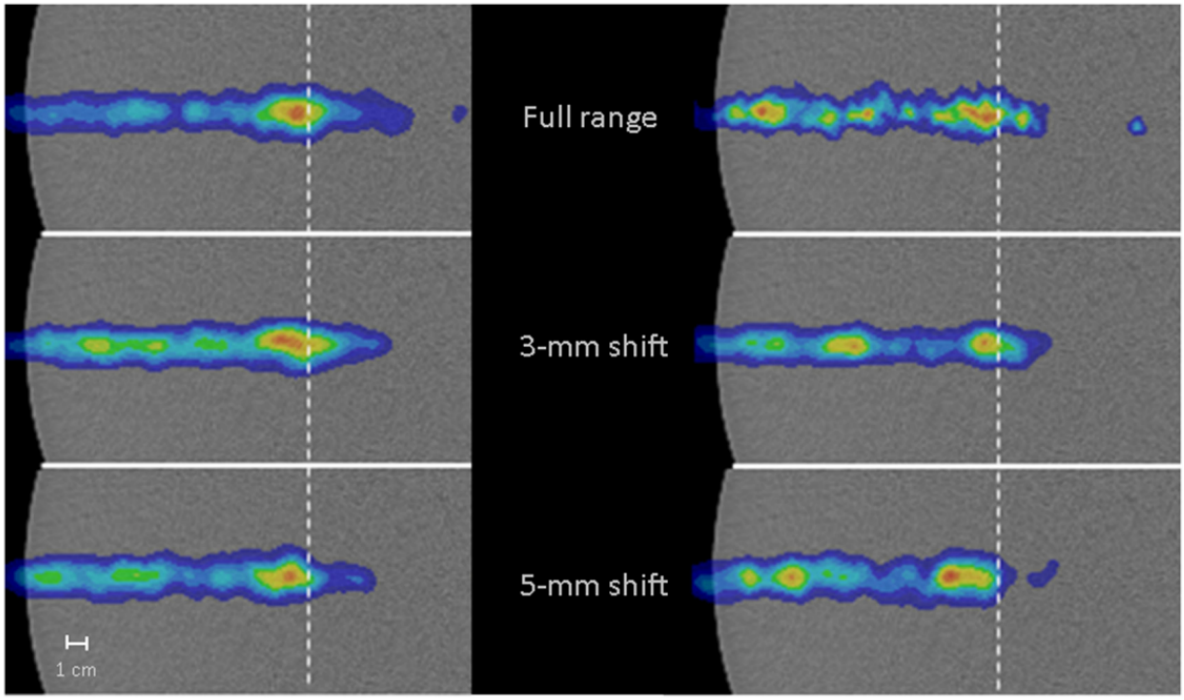


Figure 3.10: PG images obtained with the Polaris CC for the delivery of 6.29×10^8 (left) and 1×10^8 (right) protons for the full range of the irradiated 120 MeV PB (top), 3 mm range shifted PBs (middle), and 5 mm range shifted PBs (bottom), each normalized to the respective maximum value. The dashed lines represent the position of the distal 90% of the full range pencil beam (Draeger et al. 2018). The corresponding profiles are shown in figure 3.11

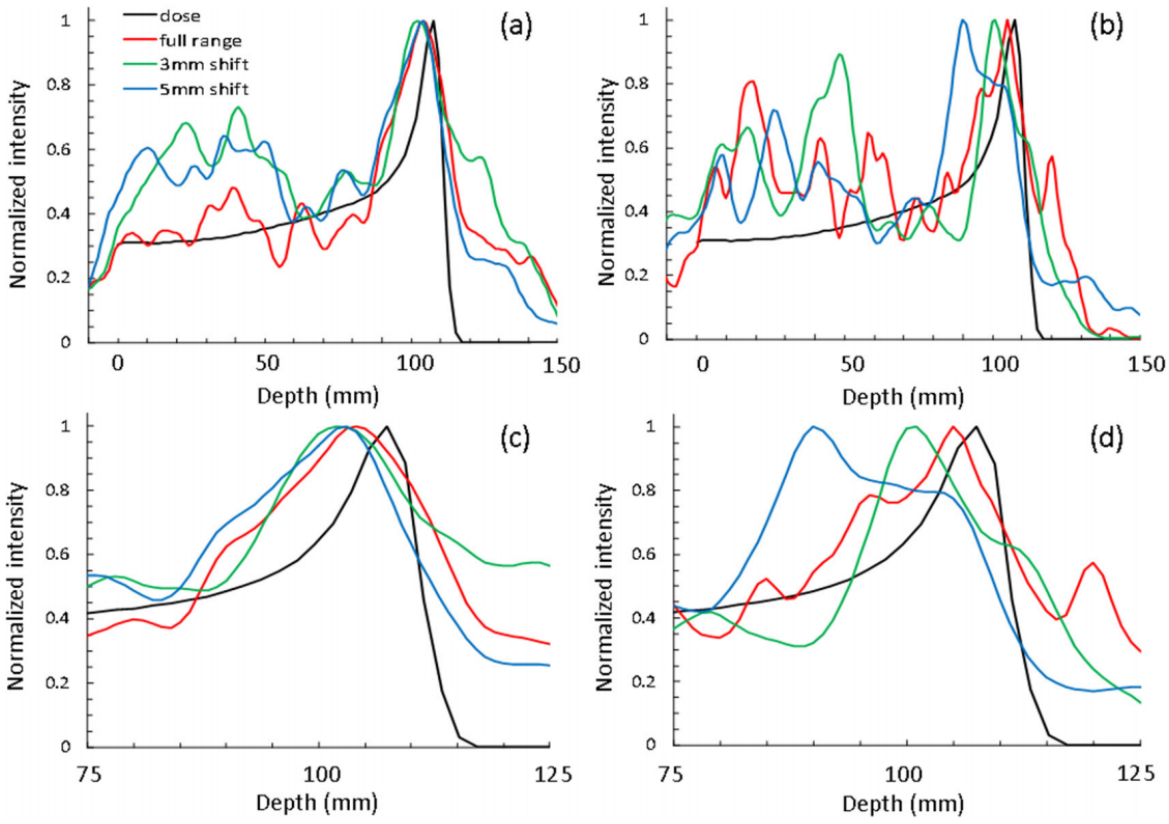


Figure 3.11: PG emission profiles extracted from the CC images of figure 3.10 for the delivery of 6.29×10^8 (a) and 1×10^8 (b) protons for the full range and for the range shifts of 3 mm and 5 mm compared to the central axis depth dose profile for the full range 120 MeV PB. (c) and (d) show a close up of the final 5 cm of the beam penetration depth to highlight the shift in the PG profile as the beam range was shifted (Draeger et al. 2018).

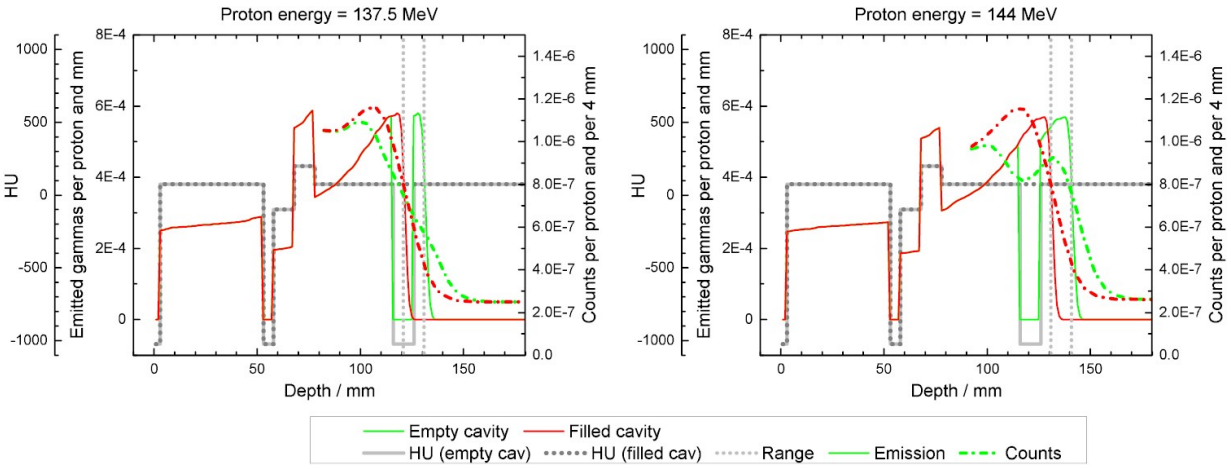


Figure 3.12: Simulation of the expected PG detection (dash-dotted) and the PG emission with energies between 3-8 MeV (solid colored lines) for the irradiation of an anthropomorphic head target with (green)/without (red) air cavity. The Hounsfield units (HU) of the irradiated target and the range of the protons are given by the gray solid and dotted lines, respectively. Left: 137.5 MeV protons with range 5 mm beyond the air cavity. Right: 144 MeV protons with range 5 mm beyond the air cavity. In the right figure, although the anatomy has changed, there is almost no PG shift detected as the algorithm calculates the shift by minimizing the area between the falloffs of both PG profiles (dash-dot) (Priegnitz et al. 2015).

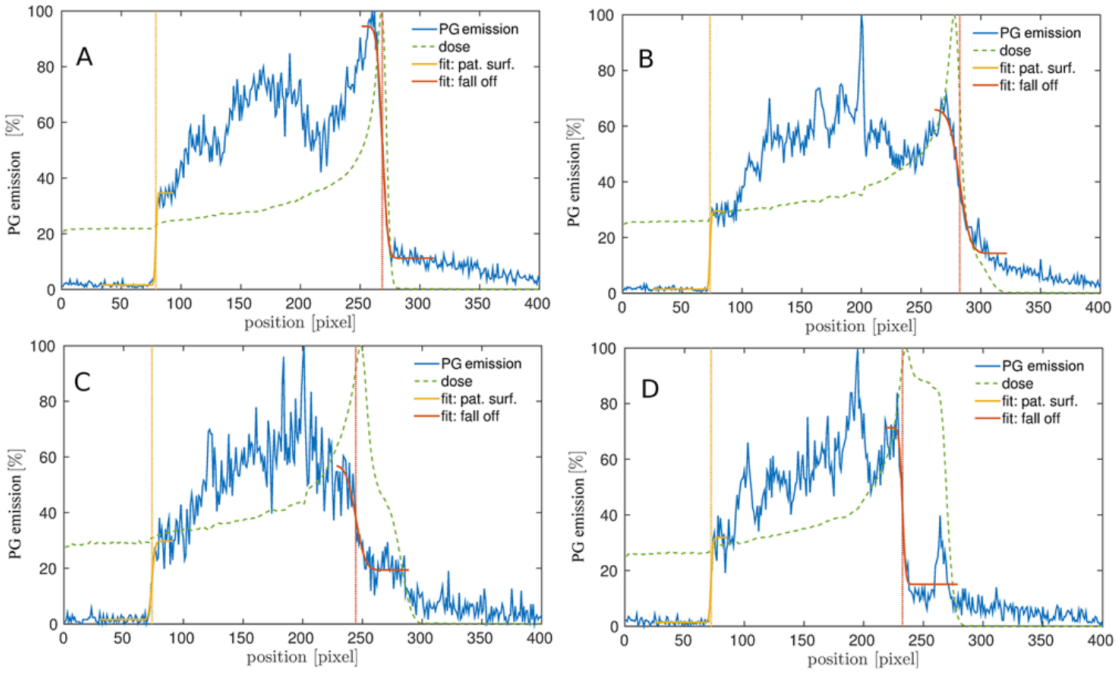


Figure 3.13: Depth dose profiles and PG emissions along with the corresponding PG falloff fittings of 4 proton PBs simulated in a clinical CT image of a prostate patient. In (B), the dose falloff is not as sharp as that in (A), resulting in a gradual PG falloff. (C) and (D) show profiles with large variation compared to a typical steep pristine Bragg peak in (A), due to distal tissue heterogeneities (Schmid et al. 2015).

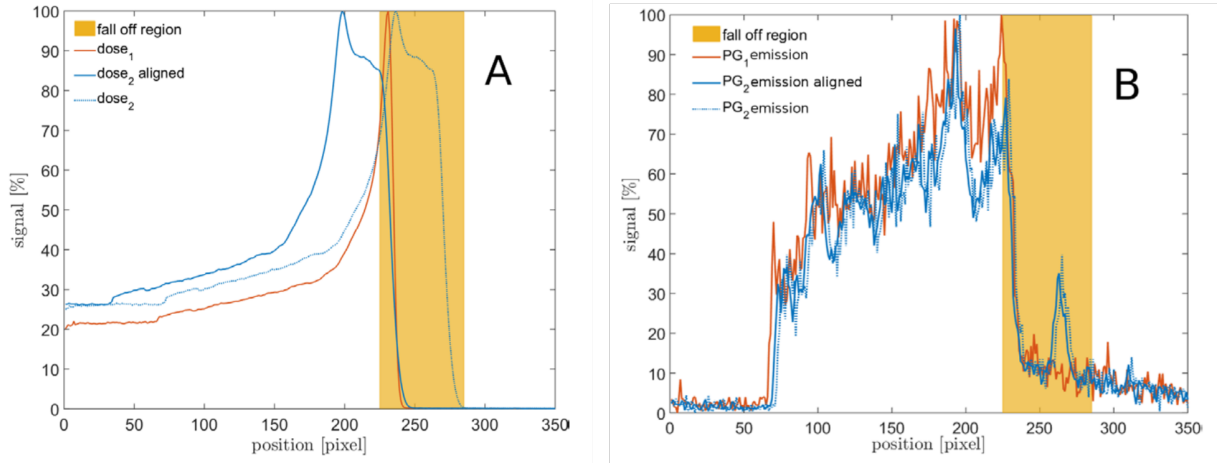


Figure 3.14: The shift of the dose and PG laterally integrated profiles of a proton PB delivered on CTs of a prostate patient at different time points are shown on the left and right, respectively. The actual dose shift cannot be correctly estimated using the PG shift. In this case, using the PG signal as a surrogate of the beam range could lead a misevaluation of the dose shift (Schmid et al. 2015).

Chapter 4

Computational Tools used

Monte Carlo simulations are used in this work to calculate the dose and PG distributions based on the patient CT. The first part of this chapter introduces the Monte Carlo treatment planning system for proton therapy along with other computational platforms used in this research, and how the dose and PG signals are scored. The second part of this chapter introduces the basic knowledge of machine learning, which is used for pattern recognition of the PG falloff used in the signal analysis.

4.1 Computational platforms for treatment planning and dose/PG distributions

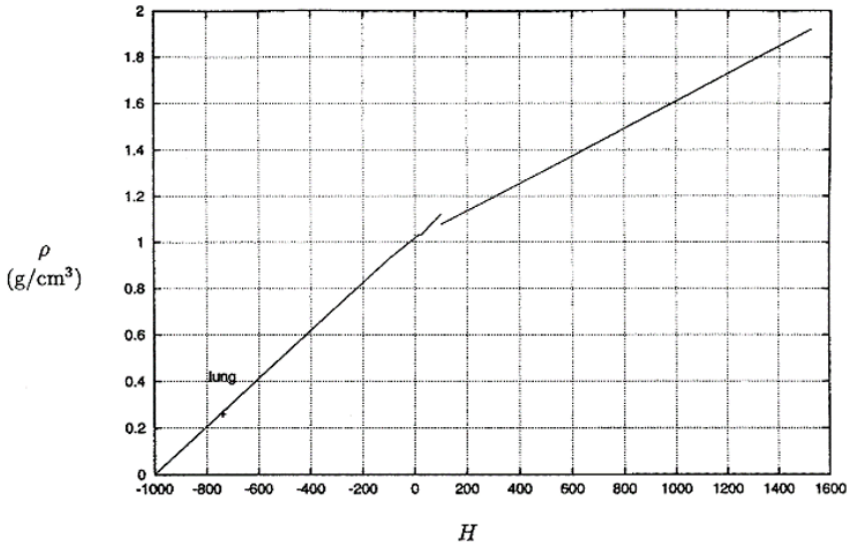
4.1.1 Patient data

In the first part of this work, the planning X-ray CT scans of three anonymized head and neck (called H&N in the following) cancer patients (H&N1, H&N2 and H&N3) are used.
The second part of this work focuses on the performance of the developed methods on different CTs of the same patient taken at different points in time. The CTs of one H&N and one prostate cancer patient (H&N4 and PROST1) at three different time points are used. The planning CT is called CT1 while the subsequent CTs are called CT2 and CT3, respectively.
For the third part of this work, the proposed approach is tested taking into account a realistic detector response, in combination with an indicator based on the PG camera position. The same CTs as in the second part of this work are used.
All the CTs refer to anonymized data of patients undergoing photon therapy at the University Hospital of the Ludwig-Maximilians-Universität München. The CT data were acquired at a Toshiba Aquilion LB scanner (Toshiba Medical Systems, the Netherlands) and the corresponding images were reconstructed on a $1.074 \times 1.074 \times 3$ mm^3 grid.

4.1.2 Monte Carlo simulations

Monte Carlo simulations are based on algorithms that repeat random samplings so that expected values of the variables of interest are obtained by taking the mean of those independent samples. This method is based on the law of large numbers. The use of MC techniques in the field of medical physics has been increasing since the 1970s (Andreo 1991, Yang & Bednarz 2013, Andreo & Benmakhlouf 2017). Different MC tools are used for MC simulations in medical physics, e.g. PENELOPE (Baro et al. 1995, Salvat & Sempau 1996, Sempau et al. 2001), MCNP (Brown & Martin 2003), Geant4 (Agostinelli et al. 2003, Allison et al. 2006, Allison et al. 2016) and FLUKA (B"ohlen et al. 2014) among others. MC simulation of the three-dimensional (3D) dose deposition, PG emission and PG detection in the first and second parts of this thesis is directly conducted using Geant4. For the last part of this work, a fast MC engine for proton therapy named MCsquare (many-core Monte Carlo) (Souris et al. 2016) is used for the 3D dose simulation and PENELOPE precaculated data are used for the fast analytical calculation of PG emission and detection (Sterpin et al. 2015).
Geant4 is a MC platform for the simulation of the passage of particles through matter. It is the successor of the Geant series of software toolkits developed by CERN using C++. Geant4 includes packages for handling geometry, tracking, detector response, run management, visualization and user interface, allowing users to start immediately on the more important aspects of the simulation while less time is needed to be spent on the low level details. In this study, the Monte Carlo simulations were performed with the version Geant4.10.02.p01 based on an extension of the DICOM example (Resch et al. 2017). The transport and interactions of all primary and secondary particles, e.g. protons, electrons, neutrons, gammas, nuclear fragments etc, are simulated using the predefined QGSP_BIC_HP Geant4 physics list. This package uses the Geant4 Binary cascade for primary protons with energies below 10 GeV, and can well describe the production of secondary particles in hadronic interactions of proton and neutrons with nuclei. The low-energy gamma and electron interaction processes including Doppler-broadening effects are covered by the electromagnetic physics list (Geant4 2018). This package and physics lists were chosen as they offers a good compromise between performance and speed. The range cut of the secondary production threshold is set to 1 mm, corresponding to the CT grid size. All settings are the same as in previous work on PG monitoring, e.g. (Schmid et al. 2015, Lozano et al. 2016).

For calculations in the patient anatomy, the patient CTs are converted to tissue composition and mass density to construct a Geant4 geometry for the TP simulation using a lookup table firstly proposed by (Schneider et al. 2000) (figure 4.1).
For scoring the quantities used in this thesis work in Geant4, the following approaches were followed.
Dose and LET$_d$ scoring: The 3D dose and LET$_d$ of the simulated irradiation is scored on the CT grid (Resch et al. 2017). At least 1×10^5 protons are needed for a reasonable statistics in the spot-by-spot dose calculation, while the particle number for the TP simulation is set to the total number of around 1×10^7, as recommended by (Resch et al. 2017).



	w_i (pp)											
H	H	C	N	O	Na	Mg	P	S	Cl	Ar	K	Ca
−1000–−950			75.5	23.2						1.3		
−950–−120	10.3	10.5	3.1	74.9	0.2		0.2	0.3	0.3		0.2	
−120–−83	11.6	68.1	0.2	19.8	0.1			0.1	0.1			
−82–−53	11.3	56.7	0.9	30.8	0.1			0.1	0.1			
−52–−23	11.0	45.8	1.5	41.1	0.1		0.1	0.2	0.2			
−22–7	10.8	35.6	2.2	50.9			0.1	0.2	0.2			
8–18	10.6	28.4	2.6	57.8			0.1	0.2	0.2		0.1	
19–80	10.3	13.4	3.0	72.3	0.2		0.2	0.2	0.2		0.2	
80–120	9.4	20.7	6.2	62.2	0.6			0.6	0.3			
120–200	9.5	45.5	2.5	35.5	0.1		2.1	0.1	0.1		0.1	4.5
200–300	8.9	42.3	2.7	36.3	0.1		3.0	0.1	0.1		0.1	6.4
300–400	8.2	39.1	2.9	37.2	0.1		3.9	0.1	0.1		0.1	8.3
400–500	7.6	36.1	3.0	38.0	0.1	0.1	4.7	0.2	0.1			10.1
500–600	7.1	33.5	3.2	38.7	0.1	0.1	5.4	0.2				11.7
600–700	6.6	31.0	3.3	39.4	0.1	0.1	6.1	0.2				13.2
700–800	6.1	28.7	3.5	40.0	0.1	0.1	6.7	0.2				14.6
800–900	5.6	26.5	3.6	40.5	0.1	0.2	7.3	0.3				15.9
900–1000	5.2	24.6	3.7	41.1	0.1	0.2	7.8	0.3				17.0
1000–1100	4.9	22.7	3.8	41.6	0.1	0.2	8.3	0.3				18.1
1100–1200	4.5	21.0	3.9	42.0	0.1	0.2	8.8	0.3				19.2
1200–1300	4.2	19.4	4.0	42.5	0.1	0.2	9.2	0.3				20.1
1300–1400	3.9	17.9	4.1	42.9	0.1	0.2	9.6	0.3				21.0
1400–1500	3.6	16.5	4.2	43.2	0.1	0.2	10.0	0.3				21.9
1500–1600	3.4	15.5	4.2	43.5	0.1	0.2	10.3	0.3				22.5

Figure 4.1: The lookup tables for the conversion of CT numbers to mass density (top) and tissue composition (bottom), as proposed by (Schneider et al. 2000).

PG emission scoring: The 3D PG emission is scored on the CT grid for the PBs of interest. Only PG with energy higher than 2.5 MeV are counted as gammas with lower energy are not strongly correlated to the proton range (Kozlovsky et al. 2002, Verburg et al. 2012, Schmid et al. 2015). As the PG production process is less frequent than the energy deposition process, 5e5 protons were found to be needed to generate enough PG emission counting.

Scoring of the phase space data of PG detection: To investigate the effect of statis-

tics on the generic (i.e., not specific to a certain detector system) PG detection, a phase space (phsp) scoring detector was set surrounding the patient CT in the Geant4 simulation, and all PG reaching the detector were recorded in terms of their incident energy, position, angle, and TOF. The energy threshold of a valid PG detection event is set to 2.5 MeV which is the same as that for PG emission. The threshold of the acceptance angle is set to 4 degree to be consistent to the previous study of (Schmid et al. 2015). The phase space data of the valid PG detection events is called PG phsp data in the following. Considering the huge computational resources needed to generate enough PG phsp data, only 20 PBs were simulated from a given TP, as presented in section 5.1. For each PB, 100 statistically independent simulations, each of 1×10^6 primary protons, were conducted in parallel. Besides, simulations using 1.35×10^8 protons are performed as this value is set to the PG detectability threshold in the first part of this work.
In addition to Geant4, calculations of dose and PG emission and detection were also performed with the REGGUI platform in the last part of the work. For the dose calculation, REGGUI integrates the Monte Carlo engine MCsquare, which was designed and optimized for the last generation of Intel Xeon processor and Intel Xeon Phi coprocessors for fast MC dose simulation (Jeffers & Reinders 2013, Wang et al. 2014, Schulz et al. 2012, Rahman 2013). Heavy charged particles, e.g. protons, deuterons and alphas are simulated using the class-II condensed history algorithm (Salvat et al. 2011, Geant4 2013). Hard ionizations, with energy losses above a user-specified threshold, are simulated individually while the soft events are regrouped in a multiple scattering theory. Cross sections from ICRU 63 (Malmer 2001) are used for the sampling of the elastic and inelastic nuclear interactions.
The MC engine used for the construction of the analytical computation of PG emission and detection [Sterpin et al., 2015] is the extension of PENELOPE (Salvat et al. 2011) to protons, called PENH (Salvat 2013, Sterpin et al. 2013, Sterpin et al. 2014). PENH simulates nuclear reactions using the cross sections from ICRU 63 (Malmer 2001) for interaction of protons with elements that have a weight contribution of more than 1% to human tissues, i.e. ^{12}C, ^{14}N, ^{16}O, ^{31}P and ^{40}Ca (Sterpin et al. 2013). Elastic and inelastic electromagnetic collisions are calculated by a mixed (class II) algorithm (Salvat et al. 2011, Salvat 2013).

4.1.3 MC TP using CERR

TPs in this work are created using the particle extension of CERR which is an open source environment for TP researches based on Matlab (version 2014).

To make a MC TP using CERR (Resch et al. 2017), the patient CT as well as the contouring of the organs are firstly imported (figure 4.2). In this work, the planning target volume (PTV) is set to the default PTV in the corresponding scenario in the initial photon plan. To reduce the computational resources needed for the TP, the OAR for the TP optimization is set to a 1 cm ring surrounding the PTV (TP OAR, an example is shown in figure 4.2, red region), instead of using the ones defined for critical structures (clinical OAR) which are used only for evaluation. The prescription dose is set to 2 and

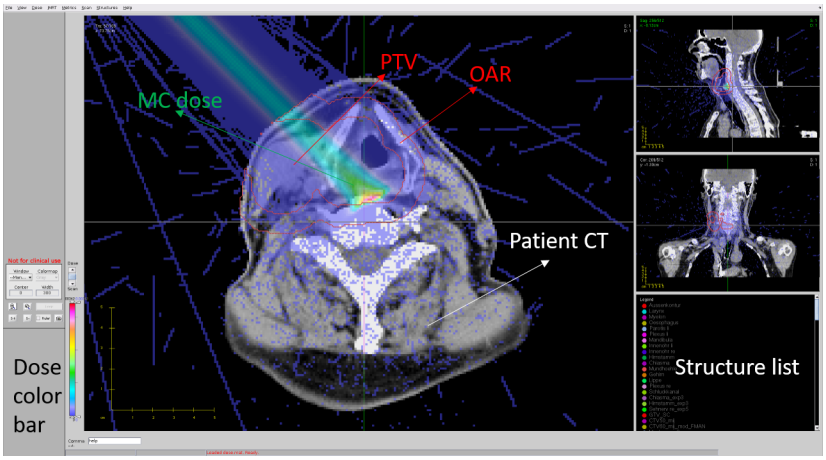


Figure 4.2: The interface of CERR TPS. The patient CT, organ contouring (in this example the PTV TP OAR of H&N1) and exemplary PB MC dose used for the MC TP optimization process are shown.

0 Gy for the PTV and TP OAR, respectively. The TP optimization (see equation 2.12) penalty factors for the PTV and TP OAR are set to 10000 and 10, respectively. Given the angle(s) of the treatment field(s), all the variables used for the TP optimization, such as PB grid (PB spacing: 5 mm), dose vectors etc., are initialized by the CERR platform. TP fields are set to two (angles of 260° and 100°), two (90° and 270°), two (45°and 270°), one (315°) and two (90° and 270°) for the patient H&N1, H&N2, H&N3, H&N4 and PROST1, respectively. Files needed for the Geant4 simulation, e.g. PBs settings and geometry are then exported by CERR. MC dose, PG emission and LET_d distribution data are then generated by Geant4 simulation for each individual PB. The generated Geant4 dose files are imported to CERR and the MC TP is created by optimizing equation 2.12 using a gradient descent algorithm.
Although in this work radiobiological considerations have been restricted to the evaluation of LET_d distributions, it is worth mentioning that CERR is also able to create TPs using biological dose based on different relative biological effectiveness (RBE) models (Resch et al. 2017).

4.1.4 Camera response simulation based on REGGUI

REGGUI is an image processing open-source platform for adaptive proton therapy based on Matlab (Janssens et al. 2009, Kurz et al. 2015). It allows the import of CT images, TPs, contours etc for further workflows, e.g. for PG simulation (both emission and detection). In particular, for the workflows developed for all investigations using multiple CT datasets of a given patient, REGGUI is firstly employed to register the CTs at different time points for the same patient, thus all the CTs for the same patient have the same isocenter and coordinate system.

For the last part of this work, addressing the response of a realistic PG detector, the

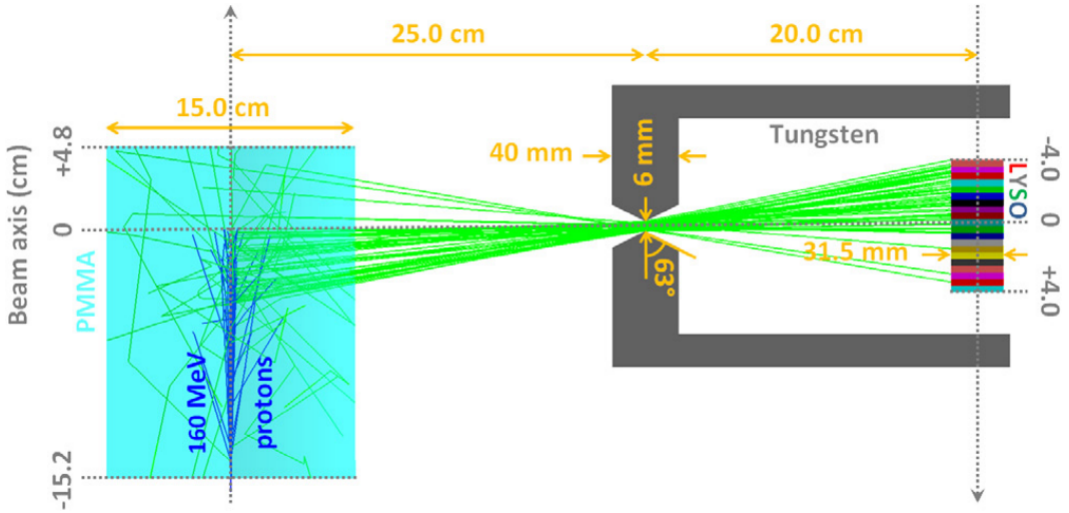


Figure 4.3: Schematic description of the PG imaging system used in REGGUI. Emitted PG are collimated by a knife-edge slit collimator and imping on LYSO crystals. The nominal magnification factor for the given geometrical distances is 0.8 (Sterpin et al. 2015).

spot-by-spot dose distribution is generated using REGGUI based on MCsquare (Souris et al. 2016). The PG emission and PG detection using the knife-edge camera (Perali et al. 2014), consisting of a photodetection system and a knife-edge slit tungsten collimator mounted on a dedicated trolley positioning system (see section 3.1.1 in chapter 3), is simulated using the REGGUI-PG package developed by (Sterpin et al. 2015). Since a full blown MC simulation is computationally expensive, the REGGUI software firstly calculates the 1D PG emission using a pre-computed MC simulation (based on the PENH extension to PENELOPE introduced in section 4.1.2) of the PG emission along the proton track as a function of incident proton energy and tissue composition. The emission information is recorded in terms of the spectrum of protons (from 3 to 8 MeV in 1 MeV steps). The PG emission is then converted to the detection using a convolution of the physical PG emission with the camera response depending on the geometry (target, PG camera position). The effects of self-absorption in the target, the camera optical effects, i.e. the magnification factor, and the interaction between high energy gammas and matter, i.e. photo effect, incoherent scattering as well as pair production are included in the convolution kernel (Sterpin et al. 2015). The simulated camera (figure 4.3) relies on 504 cm^3 of lutetium-yttrium oxyorthosilicate scintillating crystals distributed over 2 rows of 20 slabs vertically aligned one on top of the other. Each slab is 31.5 mm thick, 100 mm high, 4 mm in width along the beam axis, and together with its mirror slab in the other row, constitutes one of the 20 bins of the 1D detection. The field-of-view (FOV) of the camera can be adjusted by modifying the distance from the center of the collimator to the center of the detector crystals and/or the distance from the center of the collimator to the beam

axis. On the beam axis, a total of 100 mm FOV is covered by the crystals. The energy detection window is set to 3-6 MeV (Perali et al. 2014, Sterpin et al. 2015, Xie et al. 2017).

4.2 Spot aggregation and TP re-optimization

Sufficient statistics is required for PG imaging. In traditional TP, only a few PBs reach such PG detectability threshold (figure 4.4).

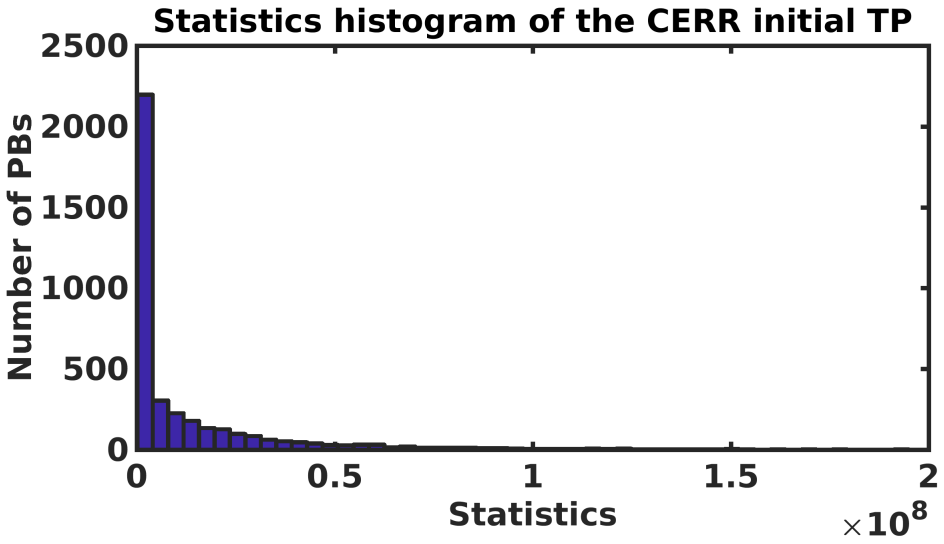


Figure 4.4: Statistics of PBs in the CERR initial TP for one of the investigated H&N cancer patients. Only 12 PBs have more than 1.35×10^8 protons among the 3811 PBs used. Hence, the majority of the PBs are not guaranteed to be detectable and to provide reliable information for PG imaging.

4.2.1 Spot aggregation

In the work of (Xie et al. 2017), the lateral spacing between PBs is 5 mm, while the individual PBs have a Gaussian-shaped lateral profile with 5-6 mm sigma, thus PBs are overlapping through part of their distribution. To achieve the required statistics for reliable PG monitoring with the used knife-edge camera, PBs nearby are aggregated (cf. section 3.2 and 5.2) at the cost of degradation in lateral spatial resolution. It has been shown that a sigma value of 4-7 mm for the Gaussian-weighted aggregation has only a limited impact on the lateral spatial resolution (reduced from 5 mm to 8.6 mm) but achieves improved signal to noise ratio. Reproducibility studies with both simulated and measured data show that precision better than 2 mm can be achieved for a large proportion of spots for aggregated counting statics exceeding 2×10^8, with only a small reduction in lateral resolution.

However, the work of (Xie et al. 2017) was conducted for a relatively homogeneous case of a brain tumor. In a more complex target with greater tissue inhomogeneities, spot aggregation will require careful spot selection and unavoidable range mixing will have to be taken into account in the data analysis (Xie et al. 2017).

4.2.2 TP re-optimization

An alternative novel approach is proposed in this work to re-optimize the TP by boosting a few selected reliable PBs which are not significantly affected by the tissue heterogeneities. More details and the comparison to the spot aggregation are described in chapter 5 and 6.

4.3 Basic concepts of machine learning

A basic machine learning (ML) model, neural network, is used in this study for the pattern recognition of PG profiles. Unlike traditional analytical approaches which are sets of explicitly programmed instructions, machine learning trains the algorithms on data inputs and use statistical method in order to output values that fall within a specific range, usually the maximum likelihood estimation. A deeper introduction to this method can be found in (Goodfellow et al. 2016), while the basic ideas are summarized in the following.

4.3.1 General idea of machine learning

In principle, the data used for the ML model optimization are divided into three independent data sets: training, validation and test dataset.
The model is initially fit on a training dataset (James et al. 2013), which is a set of examples used to adjust the parameters of the model. Successively, the validation dataset is provided to the fitted model, thus the model is evaluated using a cost function. The purpose of using independent datasets for fitting and evaluation is to avoid overfitting (see figure 4.5). The ML model is then kept fitted using training dataset until the cost calculated using the validation dataset has been minimized. Finally, the fully trained model is output and can be used for the prediction of the new data. An unbiased evaluation of the final model is given by the cost function calculated using the test dataset. The workflow is shown in figure 4.6.
Depending on the form of the data provided for the training, ML algorithms are classified into mainly supervised learning and unsupervised learning.

Supervised learning
The input data provided to the machine learning are labeled with the desired output. The idea is to train the model to evolve by minimizing the difference between its prediction (or guess) and the actual output until this difference is acceptable. The fully trained model can be then used to predict the label of

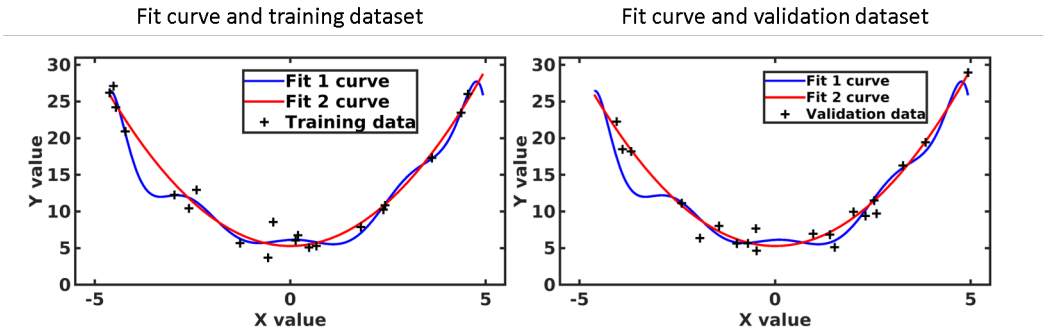


Figure 4.5: An exemplary training set and validation set from the same statistical population are shown as black spots. Two predictive models (blue and red) are fitted to the training dataset (spots shown on the left). The sum of squared errors (SSE) between the training dataset and the fit 1 / fit 2 are 21.5 / 34.6, respectively. This means that if the cost function was calculated using training data, fit 1 would be identified as being better. The SSE between the validation dataset (spots on the right) and the fit 1 / fit 2 are 101.2 / 28.6, respectively. Thus the fit 2 would be favoured by this approach. The red curve overfits the training data much less since it matches both training and validation datasets

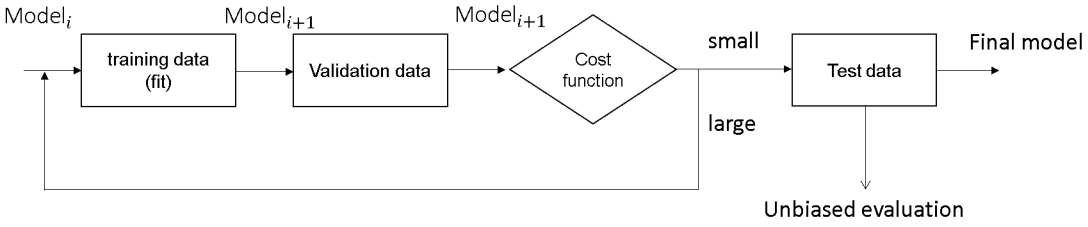


Figure 4.6: Schematic workflow of machine learning training. The model is only changed in the training data (fit) step.

new unlabeled input. This type of machine learning consists of reggression and classification (Alpaydin 2010).

Unsupervised learning

The input data provided to the model are not labeled. The algorithm is dedicated to search potential patterns in datasets without pre-existing labels. It is also known as self-organization and allows modeling probability densities of given inputs (Hinton & Sejnowski 1999). Two of the main approaches are principal component (Pearson 1901, Hotelling 1933) and cluster analysis (Garfinkel-Castro et al. 2020).

4.3.2 Introduction of neural network algorithms

This work uses only supervised neural network learning for classification (figure 4.7).
Neural network learning (NNL) is used to extract potential logic behind a group of known

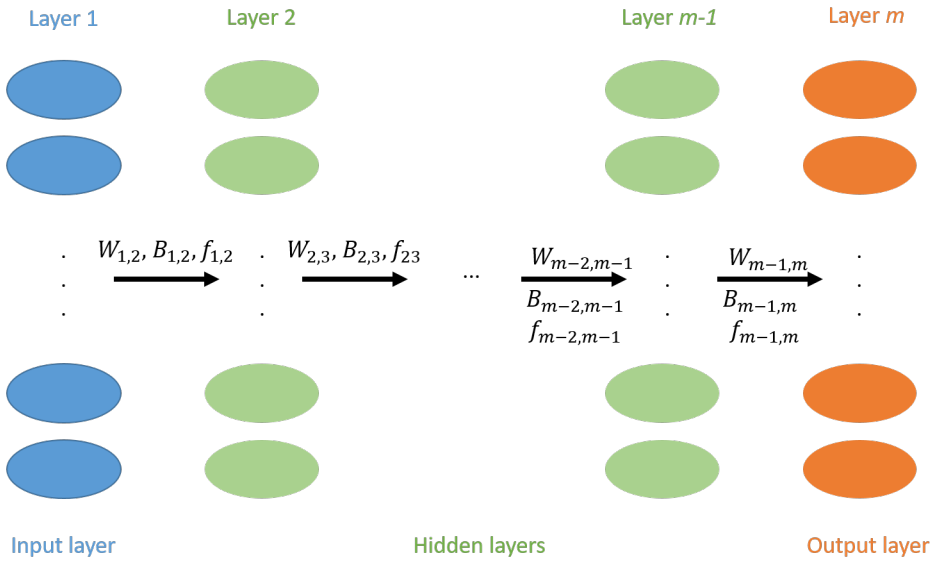


Figure 4.7: Schematic structure of neural network learning. The sizes of the input layer and output layer are the same as the sizes of input data and the label, respectively. The sizes and number of the hidden layers are assigned by the user.

data A_i ($a_{i,1}$, $a_{i,2}$, ..., $a_{i,n}$) and known label B_i ($b_{i,1}$,$b_{i,2}$, ..., $b_{i,m}$), and then predict the result Y (y_1, y_2, ..., y_m) of new unlabeled input X (x_1, x_2, ..., x_n) under the same logic. The basic neural network used in this study, also called artificial network, consists of one input layer (one vector), hidden layers (one or more vectors), and one output layer (one vector). The lth vector V_l and the $(l+1)$th vector V_{l+1} are linked by a matrix $W_{l,l+1}$, and a bias vector $B_{l,l+1}$ as well as a nonlinear function (called activation function) $f_{l,l+1}$:

$$V_{l+1} = f_{l,l+1}(V'_{l+1}) \tag{4.1}$$

in which V'_{l+1} equals the dot product of $W_{l,l+1}$ and V_l plus the bias vector $B_{l,l+1}$:

$$V'_{l+1} = W_{l,l+1} \cdot V_l + B_{l,l+1} \tag{4.2}$$

There are different types of activation functions such as sigmoid ($f(x) = 1/(1+e^{-x})$), tanh ($f(x) = tanh(x)$), ReLu ($f(x) = max(0, x)$) etc. The selection of the activation depends on the task of the machine learning and a tradeoff between efficiency and accuracy has to be made. The ReLu function is a general activation function and is used more and more nowadays.
The logic behind the known input A_i and label B_i is discovered and stored in the model

structure along with all those weight matrices and bias vectors via the training process. The fully trained model is then used to predict the label of the new input in the application process. To create a NNL model, the NNL structures (e.g. activation function types, size and number of the relevant layers etc) is firstly assigned by the user. The size of the input layer and output layer is set to the same size as the input data and the label, respectively. The number and sizes of the hidden layers are assigned by the user, depending on the task and the performance of the model. The values in the weight matrices and bias layer vectors are initialized randomly (usually from 0 to 1). The input layer L_1 is set to the example data A_i and all the values of other layers are calculated using function 4.1 and 4.2. The output layer, called O_i, is thus obtained and is compared to the label B_i using a loss function:

$$LOSS = LossFun(\vec{O}, \vec{B}) \tag{4.3}$$

in which

$$O_i = NNL(A_i) \tag{4.4}$$

The most commonly used loss function is the sum of squared errors function:

$$LossFun(\vec{O}, \vec{B}) = \sum_i (O_i - B_i)^2 \tag{4.5}$$

where O_i is calculated by a batch of inputs randomly chosen from the training dataset and B_i are the corresponding labels.
The model is then optimized by changing the values of the weight matrices and bias vectors using a gradient descent algorithm:

$$M_{i+1} = M_i - \lambda \frac{\partial loss}{\partial M_i} \tag{4.6}$$

in which λ is the step size of the optimization, assigned by the user.
The data from the input dataset are kept feeding to the optimization until the loss calculated using the evaluation dataset meets the stopping criterium set by the user.

Chapter 5

Introducing an innovative treatment planning concept including PG monitoring

The statistics of the PB will affect the precision of PG due to fluctuations of the signal and random noise. Enough statistics is needed to generate smooth PG profiles for proton range verification. Besides, in ideal cases when a PB is delivered to a homogeneous target, the distance between the falloff of the laterally integrated PG and dose profiles is a small value, thus it is possible to infer the dose falloff from the PG falloff. However, in real clinical scenarios this conformity might be deteriorated due to e.g. tissue heterogeneities. The degree of conformity (called PG-dose correlation in the following) between the dose and PG falloffs, and the robustness to tissue heterogeneities are quantified in this thesis. The main idea of this work is to identity PBs which can carry useful information for proton range verification and provide them with enough statistical weight for reliable PG imaging, while not deteriorating the dosimetric quality of the TP.

5.1 Effect of the PB statistics on the precision of the PG signal falloff

Profiles of PGs at emission and detection with different statistics are generated by summing a different amount of independent MC simulations (cf. section 4.1.2). For example, the profiles of 10^7 primary protons are generated by summing up the results of 10 independent simulations of 10^6 protons each, thus 10 profiles of 10^7 protons can be generated using the 100 independent simulations. The profiles of PG emission and detection with different statistics are compared in figure 5.1. It can be seen that the PG detection data are highly sensitive to the statistics of the incident protons.

The precision of the PG distal fall-off signal, defined as the standard deviation of the falloffs of the PG detection profiles generated by independent MC simulations for the same PB, is shown in figure 5.2 as a function of primary protons. A precision of 1 mm can be

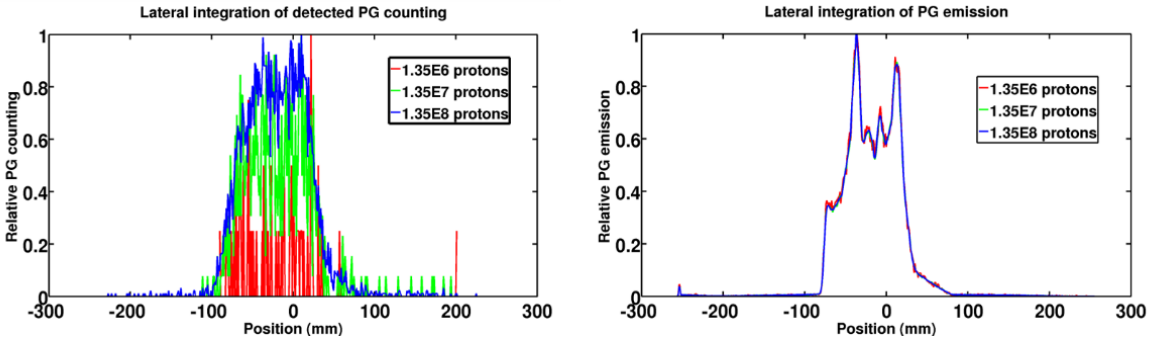


Figure 5.1: Laterally integrated depth profiles of the PG detection and PG emission distributions are shown on the left and right panel, respectively.

achieved when the number of protons of an individual PB is higher than 1.35×10^8. Only a few PBs can reach such high statistics in a traditional TP cf. figure 4.4 in section 4.2.

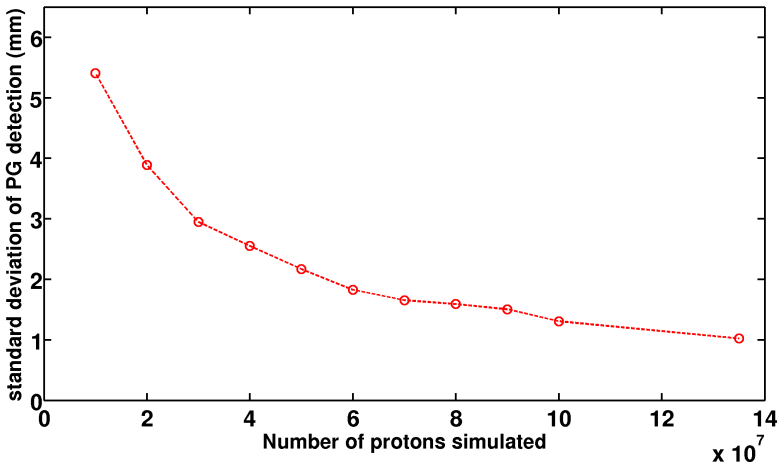


Figure 5.2: Standard deviation of the PG distal falloff based on PG detection data from different number of incident protons and statistically independent MC simulations. (Tian et al. 2018)

5.2 TP re-optimization to achieve sufficient statistics for PG imaging

CERR uses function 2.12 to optimize the TP. To boost the statistics of selected PBs above the statistical threshold w_{boost}, an additional particle number lower bound of w_{boost} is set

for the selected PBs, while 0.5 (in unit of 10^6 protons, cf. section 2.2) is set for the other PBs (considering the detectability threshold and the typical sensitivity of beam monitor systems in pencil beam scanning (Hernandez 2017)).
As a comparison, the alternative approach of PB aggregation proposed by (Xie et al. 2017) has been investigated in this work using the PB statistics from the initial CERR TP. The Gaussian kernel, normalized such that the weight of the central PB is one, was set to have a standard deviation of 7 mm, as used in the work of (Xie et al. 2017). The same PBs selected in the TP re-optimization are aggregated using corresponding nearby PBs within the range of 2 sigma, i.e. 14 mm:

$$PG_{aggregation} = \sum_i w_i \cdot \phi_i \cdot PG_i \tag{5.1}$$

where w_i is the weight of the nearby PB i in the initial CERR TP, PG_i is the corresponding PG emission profile obtained from MC simulation (where the same statistics of 5×10^5 primary protons is used for every simulated PB) and ϕ_i is the Gaussian weight of the i-th PB.

5.3 PB selection for boosting accounting for the conformities between PG and dose signal

5.3.1 Quantification of the conformities between PG and dose signal

Figure 5.3 shows an example where the PG-dose correlation is deteriorated by the tissue heterogeneities, meaning that the distal falloffs of the dose and PG laterally integrated depth profiles do not match well.

There are two approaches to analyze the correlation between the laterally integrated profiles of the PG (emission) and dose, which requires the identification of the falloff relative or absolute position: shifting (Knopf et al. 2009, Helmbrecht et al. 2013, Frey et al. 2014) and fitting (Janssen et al. 2014). The shifting method, requiring at least two profiles of the same quantity to be compared, is used in this work to determine the shift of the profiles of dose and PG of the PBs on different CTs. The fitting method, applicable to a single PG profile, tries to recognize and analytically describe the falloff pattern of the PG signal. The PG-dose correlation is identified by comparing the falloffs of the dose and PG. The first part of this work focuses on the PG-dose correlation using only CT1, thus the fitting method is applied. The shifting method will be used in the next part of this work, where the PG and dose profiles on CTs at different time points are compared.
The dose falloff position is defined by the 80% distal falloff of the laterally integrated depth dose profile (Paganetti 2018). An ideal PG signal falloff, e.g. occurring when the PB is delivered to a relatively homogeneous tissue, consists of a distal gradual tail and a sharp

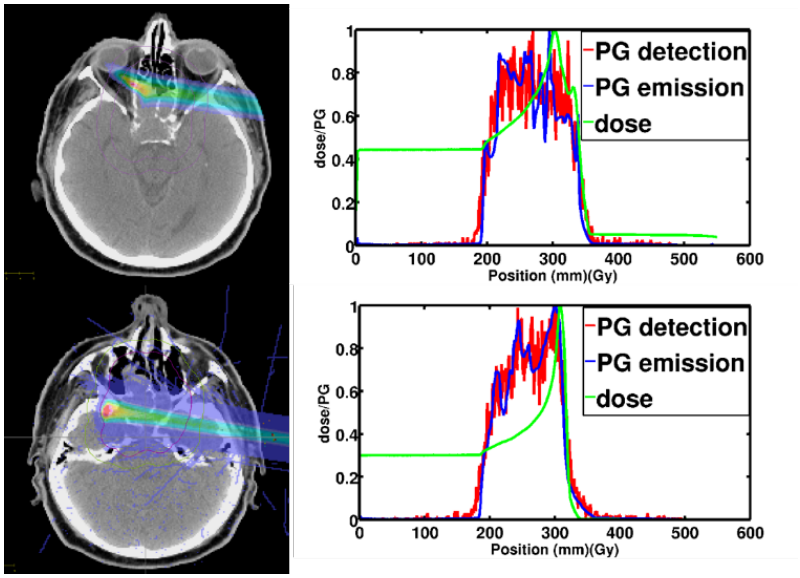


Figure 5.3: Dose distribution (in colourwash) of two PBs in the patient X-ray CT (gray scale) are shown on the left, while the laterally integrated profiles of dose, PG emission and ideal PG detection of the corresponding PBs are shown on the right. The PB on the top is affected by tissue inhomogeneities, such that the Bragg peak is distorted and the correlation between the falloffs of the dose and PG is poor. For the PB on the bottom, the falloffs of the PG profiles, both emission and detection, well correspond to the falloff of dose, thus it is possible to verify the dose falloff from the PG signal.

falloff connecting the tail and the main part of the PG profile (figure 5.4.A). A sigmoid function:

$$f_0(x) = a + (1 - a) \cdot erf[b \cdot (x - c)] \tag{5.2}$$

is employed to fit the falloff of the PG profile (Janssen et al. 2014). The position of the PG falloff is defined by the center of the fitting curve, c. In the case of a relatively homogeneous tissue (figure 5.4 A), the tail makes little contribution, thus the fitting describes the falloff well. However, due to tissue heterogeneities near the Bragg peak, the PG falloff could be more complex, affecting the PG-dose correlation (figure 5.4 B, C and D).
Three criteria are proposed for the quantification of the PG-dose correlation:

1. The distance between the falloffs of PG and dose. In the case of homogeneous tissue, this distance is supposed to be a fixed small value such that the proton range, i.e. dose falloff, could be inferred knowing the PG falloff. Let R_i be the distance between the falloff of the dose and PG of PB i. This criterion is defined by $\Delta R_i = |R_i - R_{mode}|$ in which R_{mode} is the mode estimation of all R_i of the PBs used in the TP.
2. The slope of the sigmoid fitting curve b_i of PB i. A gradual falloff means that the energy deposition is not concentrated in space due to low density medium

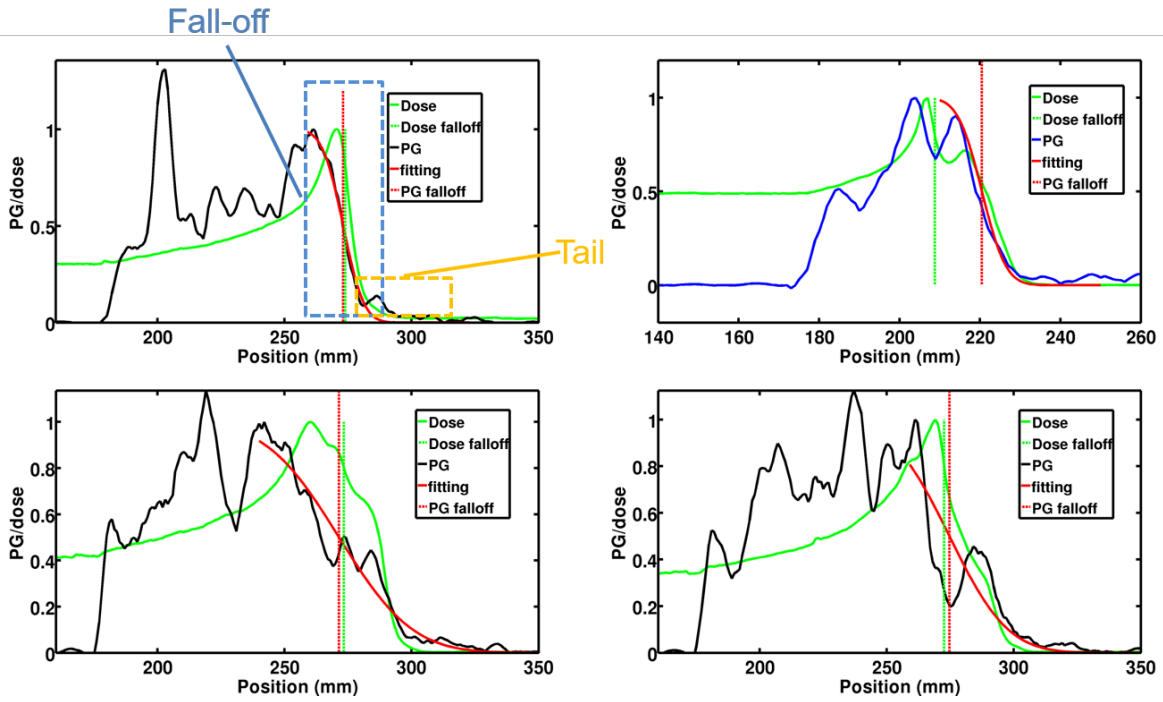


Figure 5.4: Typical examples of laterally integrated PB profiles of the PG and dose distributions calculated in a CT of a head and neck tumour patient to demonstrate the criteria quantifying the PG-dose correlation. A) The PB is delivered to a relatively homogeneous tissue. The fitting curve has a sharp falloff and the PG falloff position is close to the dose falloff. B) The distance between the falloffs of the PG and dose is large. C) The slope of the fitting is gradual, leading to a potentially unreliable PG-dose correlation. D) The sum of squared errors (SSE) between the fitting and PG is large, due to a distal peak caused by tissue heterogeneities.

or tissue heterogeneities, thus the R_i of this PB is not reliable.
3. The SSE value between fitting and the PG falloff. This criterion describes the reliability of the fitting method for the given PB.

5.3.2 PG falloff pattern recognition

To properly fit the falloff of the PG profile, the *falloff region* has firstly to be recognized. Considering the complexity of PG profiles in the case of a head and neck cancer patient, a pattern recognition algorithm is developed for the falloff identification of PG, unlike traditional fitting methods. Examples are shown in figure 5.5.
First of all, the PG profile is smoothed and linearly normalized to a maximum value of 1. The *region of interest* is defined from the most distal peak which is above 0.8 (called X_{start})

to the most distal point above 0.05 (called X_{end}). A new fitting function $f_1(x)$ consisting of a sigmoid function (see formula 5.2) $f_s(x)$ and a distal Gaussian function $f_G(x)$ is used to fit the *region of interest*:

$$f_1(x) = f_s(x) + f_G(x) \tag{5.3}$$

in which

$$f_s(x) = a + (1-a) \cdot erf[b \cdot (x-c)] \tag{5.4}$$

$$f_G(x) = d \cdot exp[-(x-e)^2/2f^2] \tag{5.5}$$

where coefficient c in $f_s(x)$ and e in $f_G(x)$ fulfill:

$$e - c > FWHM/2 \tag{5.6}$$

thus the Gaussian peak is always distal and a proper distance is guaranteed. The FWHM is the full width at half maximum of the Gaussian function $f_G(x)$:

$$FWHM = 2.35482 \cdot f \tag{5.7}$$

Combining the formulas (5.3)-(5.7), the final fitting function for the *region of interest* is:

$$f_1(x) = a + (1-a) \cdot erf[b - (x-c)] + d \cdot exp[-\frac{2.35482^2 \cdot (x-c-\varepsilon)^2}{8(\varepsilon-\lambda)^2}] \tag{5.8}$$

with constraints $\varepsilon > 0$ and $\lambda > 0$. Here the coefficients e and f are calculated by ε and λ to simplify the coding script. The fitting $f_1(x)$ as well as the sigmoid $f_s(x)$ and Gaussian $f_G(x)$ components are shown in the top raw of figure 5.5 in red and blue, respectively.
Two falloffs are provided by the sigmoid and Gaussian components of $f_1(x)$ fitting. Falloff1 is given by the sigmoid function, starting from X_{start} to the lowest point X_1 in $[X_{start}, c+\varepsilon]$. The falloff2 defined by the Gaussian formula starts from the highest point in $[X_1, X_{end}]$ to the end of the *region of interest* X_{end}. Either falloff1 $[X_{start}, X_1]$ or falloff2 $[X_2, X_{end}]$ is possible to be the true falloff of the entire PG profile which best matches the dose falloff (see middle row in figure 5.5).

Identifying the *falloff region* of PG is simple when the dose profile is available, i.e. at the planning stage:

$$F_{select}(PG_{plan}, dose_{plan}) = \begin{cases} 0, & \text{falloff1 is closer to the dose falloff} \\ 1, & \text{falloff2 is closer to the dose falloff} \end{cases} \tag{5.9}$$

Note that the dose profile is assumed to be unknown for the measurement data, i.e. when the PG is used to infer the dose falloff (MC recalculation in this work, where the PG profiles are called PG_{rec}). A supervised neural network (cf. chapter 4) learning model is built for the falloff identification in this case:

$$F_{select}(PG_{rec}) = NNL_{app}(PG_{rec}, W, B) \tag{5.10}$$

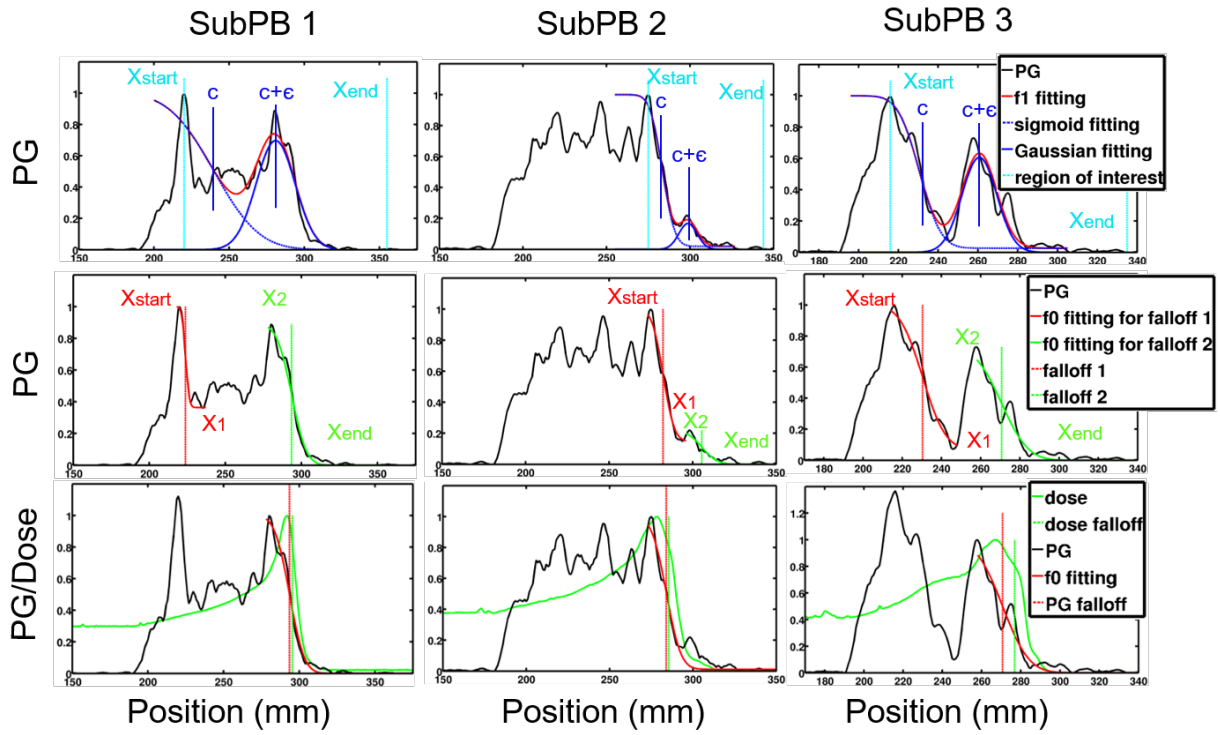


Figure 5.5: Three examples of how the pattern recognition based fitting method works. f_1 fitting is shown on the top row. The identification of the *falloff region* is demonstrated in the middle row. In the last step, the sigmoid fitting $f_0(x)$ is applied to the *falloff region* as shown by the bottom row.

in which the NNL_{app} is the neural network model while W and B are the weight matrices and bias vectors trained using the data generated in the planning stage (labeled using function 5.9):

$$\begin{cases} W \\ B \end{cases} = NNL_{train}(PG_{plan}, dose_{plan}) \tag{5.11}$$

The input vector consists of: 1) The heights of both falloffs; 2) The slopes of both falloffs; 3) the distance between the Gaussian and sigmoid functions; 4) whether the falloff1 is proximal to the weighted average position $[\sum x \cdot PG(x)]/\sum x$ of the entire PG profile (boolean), in which x and $PG(x)$ are the position in depth and PG value.
Two hidden layers (the first one has 14 arguments and the second one has 6 arguments) are employed in this work. All the activation functions are set to the sigmoid function ($f(x) = 1/[1+exp(-x)]$). The logic of this NNL is simple: a falloff with steep slope and pronounced height is preferred. Similar settings of the hidden layer, e.g. two hidden layers of 10 and 8 arguments, do not significantly affect the performance of the NNL. However, since this is a very simple NNL model based on rather intuitive criteria, the size and numbers of

the hidden layers should not be too large so that potential optimization problems e.g. in gradient descent (Ruder 2016) and exploding gradient (Pascanu et al. 2012) could be avoided.

The output layer has only one argument. Falloff2 is selected when the output is above

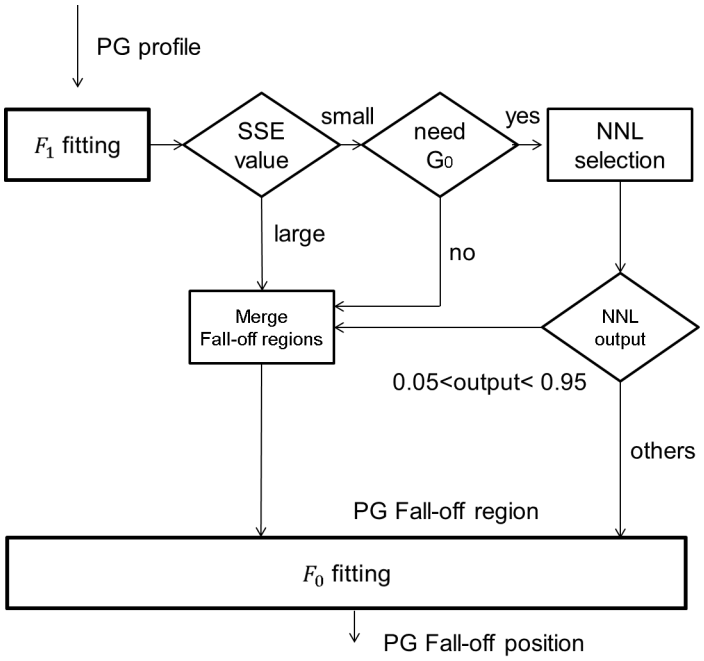


Figure 5.6: Workflow of the pattern recognition based fitting method. (Tian et al. 2018)

0.95 while falloff1 is selected when the output is lower than 0.05. For some exception cases listed below, the falloff1 and falloff2 are merged and the *falloff region* is set to the combination $[X_{start}, X_{end}]$. The PG-dose correlation degree can be easily identified by the slope and ΔR_i in those cases:

1. The Gaussian function is not necessary for the fitting of the *region of interest*, characterized by: (a) overlapping of falloff1 and falloff2, (b) a very low Gaussian peak, (c) a short distance between the Gaussian and the sigmoid function, and (d) a narrow Gaussian peak.
2. $f_1(x)$ fitting fails, showing a large SSE value (top 10% of the highest SSE values). The initial idea of introducing an additional Gaussian function is to better fit the potential distal peak affected by relatively simple tissue heterogeneities. This approach would fail when the tissue heterogeneities are more complex, i.e. the PG-dose correlation is poor. In this case the combination of falloff1 and falloff2 is regarded as the *falloff region* and the corresponding PB can be easily identified as poor PG-dose correlation in the following process.
3. An output of NNL between 0.05 and 0.95 suggests that NNL cannot select a dominant falloff among the competitive falloff1 and falloff2.

Finally, the conventional sigmoid function $f_0(x)$ is applied once the $falloff$ region has been identified and the falloff position is given by the parameter c. The exemplary algorithm and the workflow are shown in figure 5.5 and 5.6 respectively.

5.3.3 PBs selection accounting for PG-dose correlation

Considering the uncertainties of the range for those boosted PBs which will deliver high dose, only PBs with maximum proton energy which have their Bragg peak in a safety region inside the PTV planning target volume (i.e., within an inner safety margin of 1.5 mm from the edge of the PTV) are evaluated in terms of PG-dose correlation. PBs with small ΔR_i (less than 2.5 mm), steep slope ($b < -1/5$) and small SSE value (not top 33% of the highest SSE values) are identified to have a good PG-dose correlation. Additionally, PBs which have an air cavity in the Bragg peak region, i.e. CT numbers lower than -700 in a region within 2 mm range of the highest dose point, are discarded. Considering that the statistics of the selected PBs would become obviously higher than that used in the initial CERR TP, PBs are divided into different geometrical groups to avoid possible Bragg peak overlap so that hot spots are avoided. In this work, five PBs per field are selected among those with good PG-dose correlation from four nonadjacent outer groups and a random inner group (see figure 5.7).

5.4 Comparison of the re-optimized TP and initial CERR TP

In this work, the Geant4 based MC TPs (see chapter 4), called in the following initial CERR TPs, for three head and neck cancer patients (H&N1, H&N2 and H&N3), each with two treatment fields, are re-optimized by boosting 5 spots with good PG-dose correlation per field. The prescription dose is set to 2.0 Gy. For each patient, the re-optimized TPs are created by setting the w_{boost} to 1.35e8. The initial CERR TPs and re-optimized TPs are recalculated on the patient CTs using Geant4 and compared in terms of dose distribution, LET_d distribution, TP robustness to range uncertainties and translational setup errors, as well as spot-by-spot PG-dose correlation.

5.4.1 Dose distribution

The 3D dose distributions along with the dose-volume histograms (DVH) of targets and clinical OAR of all TPs created are compared in figure 5.8 and 5.9. The dose delivered by the initial CERR TPs and the re-optimized TPs is comparable. The statistics of the boosted PBs is above the set w_{boost} (1.35×10^8) and the corresponding Bragg peaks are in the PTV and near the PTV margin. The maximum dose delivered by the boosted PBs in

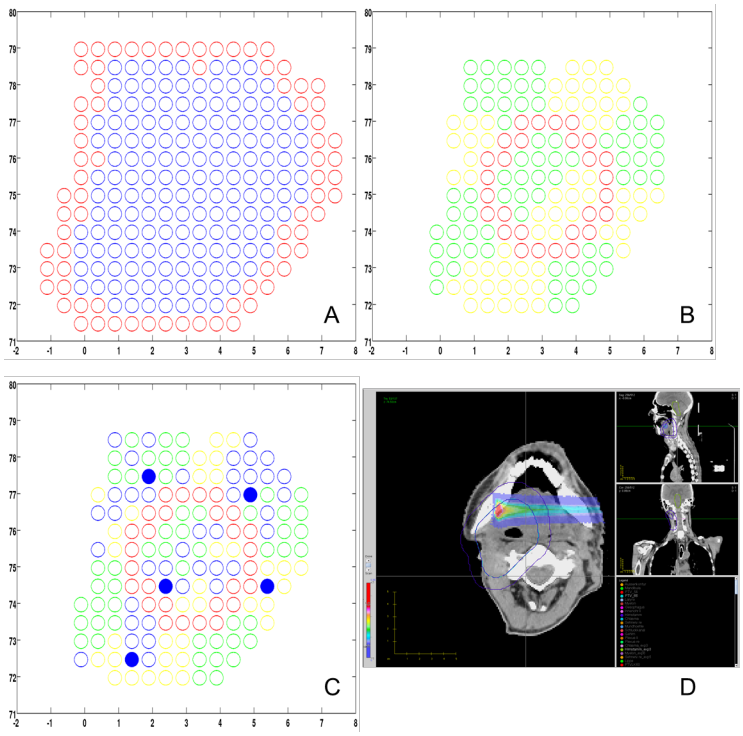


Figure 5.7: The PBs selection for boosting. A) Beams-eye-view (BEV) of the TP. The blue spots are the maximum energy PBs which have Bragg peaks inside the selected safety region within the PTV, while the red spots refer to the PBs covering the edge of the target, whose Bragg peak is outside the PTV on the given CT. B) The PBs available (blue spots in (A)) are divided into 10 equivalent geometrical groups (yellow and green spots). The red ring is a margin which separates the inner and outer zones. C) PBs are evaluated in terms of PG-dose correlation. The blue spots are presented to the user for selection. The solid blue spots are the PBs selected and boosted in this re-optimized TP. D) The dose distribution of a given PB can be shown if needed. (Tian et al. 2018)

the re-optimized TP is 0.59, 0.61, and 0.88 Gy for the three patients.

5.4.2 TP robustness

The robustness of the initial CERR TPs and re-optimized TPs are evaluated on patient 1 using MC simulations considering (1) range uncertainties of ca. ±3% by changing the CT numbers by ±3% and (2) setup uncertainty of ±3 mm by shifting the patients CT image along the x, y, z axis respectively. Figure 5.10 shows the DVHs of the TPs simulated under different uncertainties. The robustness of both plans is found to be comparable, with DVH curves of the targets overlapping and comparable trend for the clinical OARs, notably with a tendency to lower dose in the re-optimized TP.

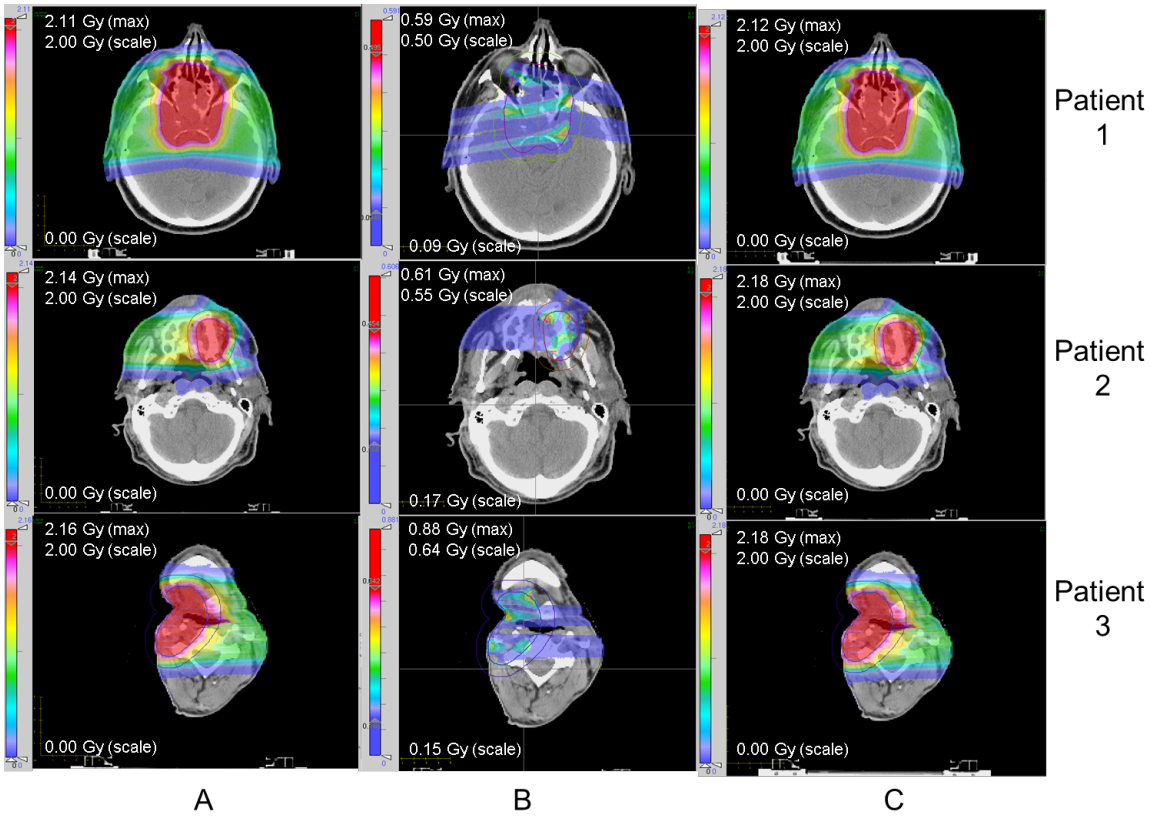


Figure 5.8: The CT scan (gray scale) with superimposed dose distribution (colourwash) of the initial CERR TPs (left column), the boosted PBs (middle column), and the re-optimized TPs with w_{boost} of 1.35×10^8 (right column) for three head and neck cancer patients. The scale and maximum dose values are shown next to the color bar. (Tian et al. 2018)

5.4.3 Dose averaged LET distribution

Considering the ability of CERR to create TPs based on different RBE models (Resch et al. 2017), which depend on the LET distribution (cf. section 2.1.1), LET_d is recorded during the MC simulation of the TPs. The initial CERR TPs and re-optimized TPs for the three considered patients are compared in terms of dose averaged LET in figure 5.11. Both TPs are found equivalent with deviations of the maximum LET_d values within approximately 2.1%.

5.4.4 Spot-by-spot PG-dose correlation and statistics

The BEV of the individual fields for the TPs created are shown in figure 5.4.4, where the spot-by-spot statistics and ΔR_i (see 5.3.1) of the corresponding PBs are illustrated by the dimension of the symbols and colour-coded. The spots with red rings are the top 5 PBs

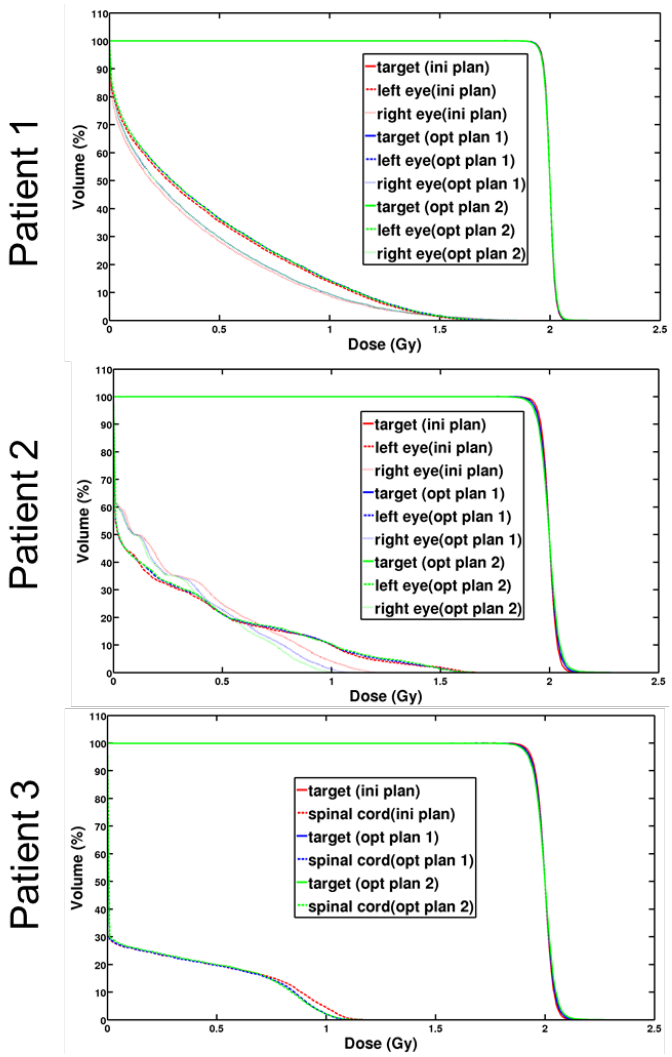


Figure 5.9: The DVHs of the initial CERR TPs (initial plan) and the re-optimized TPs (opt plan) are compared. (Tian et al. 2018)

with the highest statistics in the field.
In the initial CERR TP, the distribution and the PG-dose correlation of the highest-statistics PBs are random and not imposed, but resulting from the optimization process. In the re-optimized TPs, a few PBs with good PG-dose correlation selected uniformly across the fields are boosted and have the desired statistics. The maximum ΔR_i of the highest-statistics spots in the initial CERR TPs is more than 10 mm while this value is lower than 3 mm in the re-optimized TP.

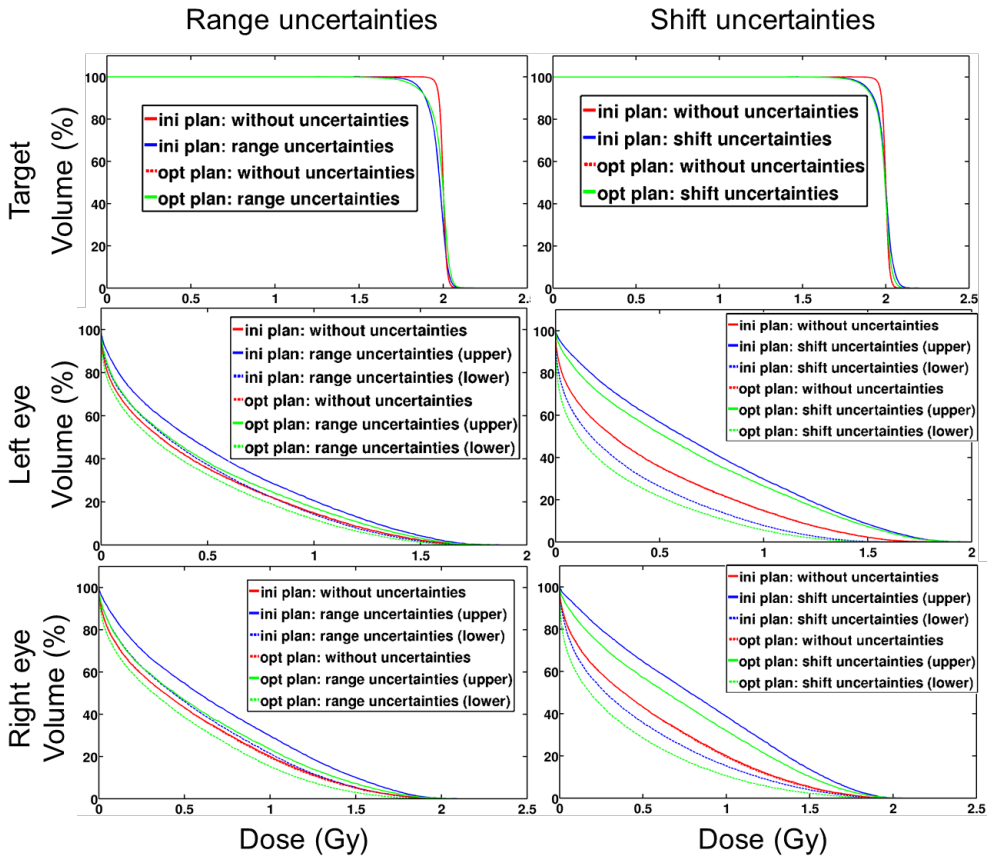


Figure 5.10: The DVHs of the initial CERR TP and re-optimized TP are compared considering both range uncertainties (left) and setup uncerntainties (right). Only the upper and lower bounds among all cases instead of all the curves are shown. Note that the solid and dashed red lines almost overlap. (Tian et al. 2018)

5.5 The robustness of the PG-dose correlation in the case of anatomical changes

In the previous study, only systematic range changes and positional (translational) errors were addressed, not accounting for the effect of unpredictable anatomical changes which can occur over the course of fractionated radiation therapy. To further investigate this effect, the dose and PG emission of PBs used in the TPs for one head and neck (H&N4, 1 field) and one prostate cancer patients (PROST1, 2 fields) are simulated on CTs acquired at three time points to evaluate the performance of the PG-dose correlation indicator, and devise an extension which better accounts for anatomical changes.

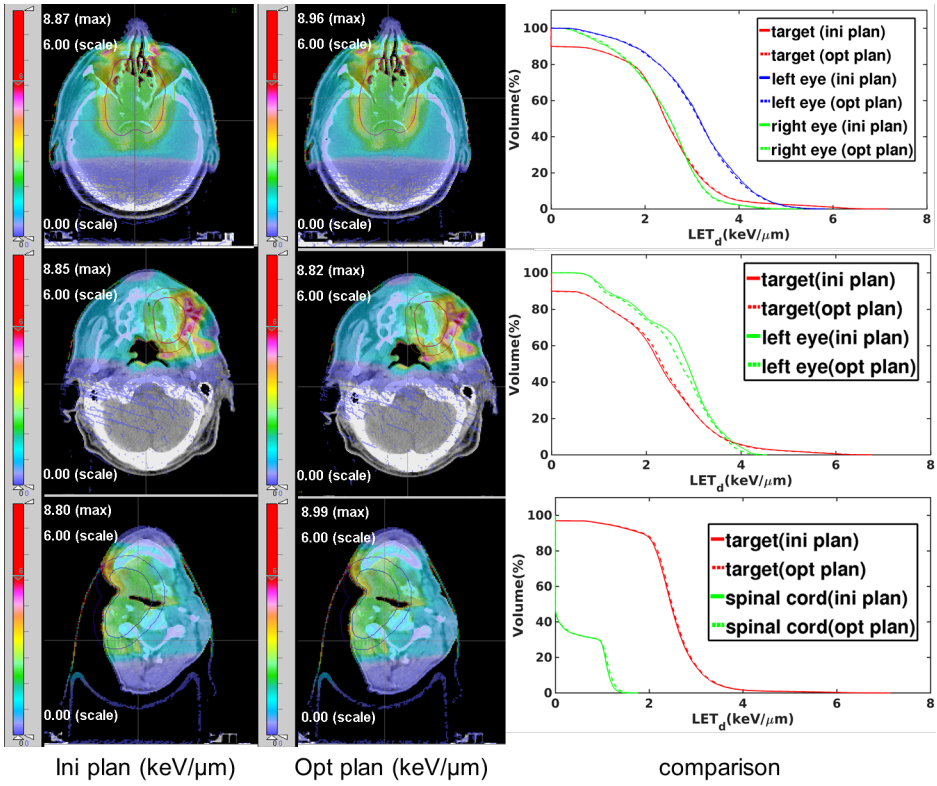


Figure 5.11: Dose average LET maps of the initial CERR TP and re-optimized TP (left) and the comparison between both TPs in terms of LET_d volume histograms (right). The initial and the re-optimized TPs are equivalent as the volume histograms overlap. The differences between the maximum dose averaged LET values of different TPs are approximately within 2.1%. (Tian et al. 2018)

5.5.1 Shifting method

To quantify the effect of anatomical changes on the PG signal, the PG emissions on different CTs of the same patient for the same PB are compared with the shifting method described in the following, and thereby the relative shift of the PG distal fall-offs is obtained. These shifts are then compared to the shifts of the fall-off of the dose profiles on the corresponding CTs, also devised from the shifting method. The difference between the shifts of the profiles of dose and PG on different CTs describes the degree of reliability of PG imaging as surrogate of the dose delivery for a given PB under consideration of anatomical changes. The shifts of PG/dose laterally integrated depth profiles are calculated by minimizing the SSE value between the profile on CT1 and the shifted (step size: 0.01 mm) profile on CT2 in a given distal fall-off region (figure 5.13). For the dose, the distal fall-off region is taken from the max value to the end of the profile. For PG profiles, the fall-off identification follows the same approach described in section 5.3.

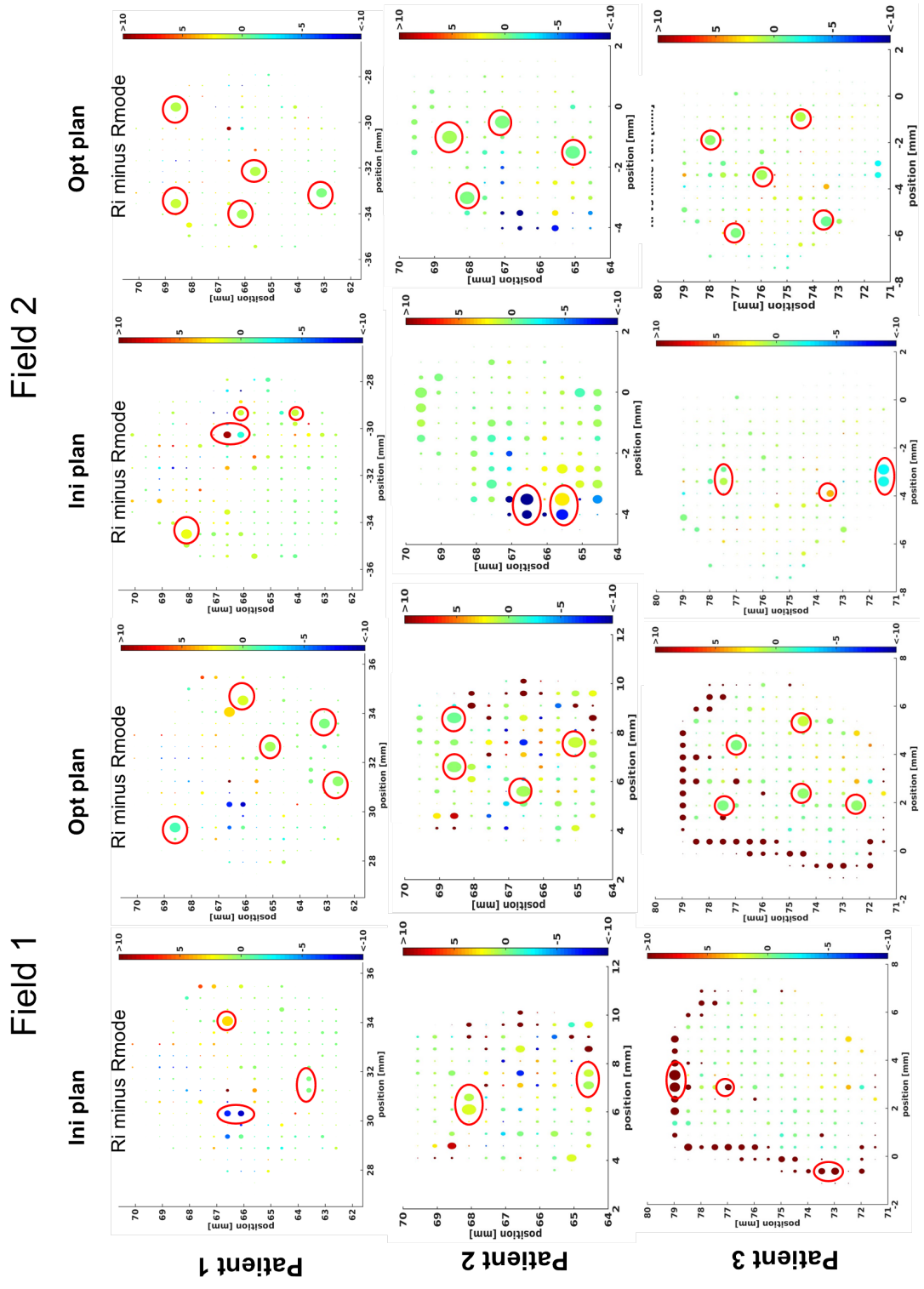


Figure 5.12: BEV of the spot-by-spot ΔR_i and statistics of the fields for the three patients. The initial CERR TPs are shown on the left and the re-optimized TPs are on the right. The ΔR_i is color coded and the radius of the spot is linearly proportional to the statistics of the corresponding PB. Spots with red circles are the 5 PBs with the highest statistics in the corresponding field. (Tian et al. 2018)

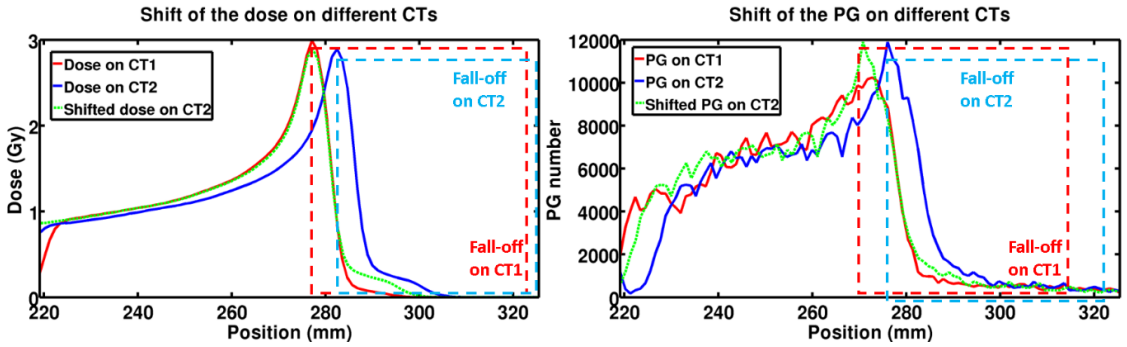


Figure 5.13: Exemplary shift calculation for the laterally integrated depth profiles of dose (left) and PG emission (right) on CT1 (red) and CT2 (blue). The shifted dose/PG profiles of CT2, which yield the minimum SSE value compared to the profiles on CT1, are shown in green. (Tian et al. 2020)

5.5.2 PG-dose correlation in presence of anatomical changes

Figure 5.14 shows the PG-dose correlation quantified as the ΔR_i (cf. section 5.3.1) on CT1 as well as the difference of the shift of the dose and PG on CT1 and CT3 for one head and neck patient (H&N1). The PG-dose correlation of a PB might change due to anatomical changes, e.g. the spot with a red circle shows good PG-dose correlation on CT1, however the PG-dose correlation on CT3 changes and thus the shifts of the dose and PG do not match.

The anatomical changes, especially for tissue in heterogeneous regions, could signifi-

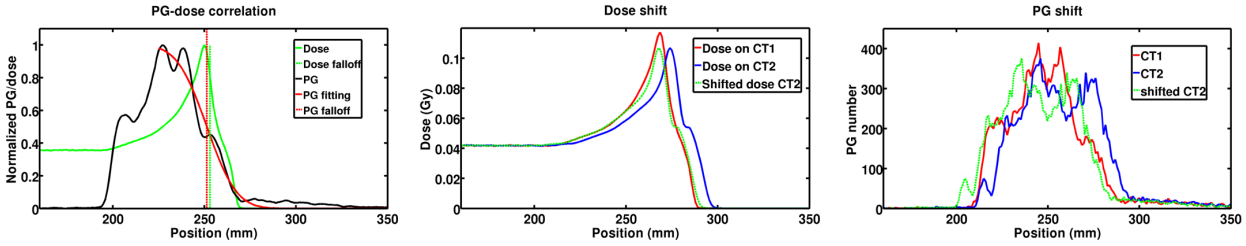


Figure 5.14: The PG-dose correlation quantified as the ΔR_i on CT1 is shown in the BEV on the left side, while the shift difference of the dose and PG on CT1 and CT3 are shown on the right. The ΔR_i and shift difference values are color coded. The details of the PG and dose profiles of the spot with the red circle are shown in figures 5.15 and 5.16

cantly change the PG-dose correlation in different CTs representing variations of patient anatomy during the course of fractionated therapy. Hence, these findings motivated the development of a new indicator to account for the effect of the tissue heterogeneities, as addressed in the next section.

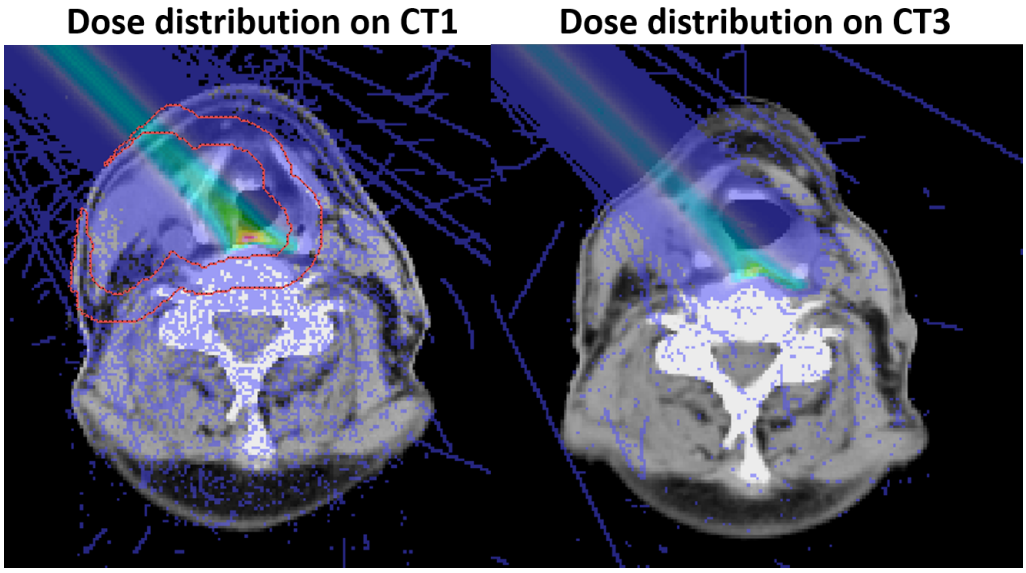


Figure 5.15: Dose distribution of the spot marked with a red circle in figure 5.14, calculated on CT1 and CT3. Note that the shape of the air cavity has changed and thus the dose distribution is affected.

5.5.3 Quantification of the effect of tissue heterogeneities

Dose surface indicator

It was proposed by (Min et al. 2014) to quantify the effect of tissue heterogeneities by using the distal 2D surface in a study of PET (cf. section 2.3.2) based range monitoring for passively scattered proton beam delivery. A similar concept is introduced to estimate the effect of tissue heterogeneities on the MC dose distribution of each individual PB in

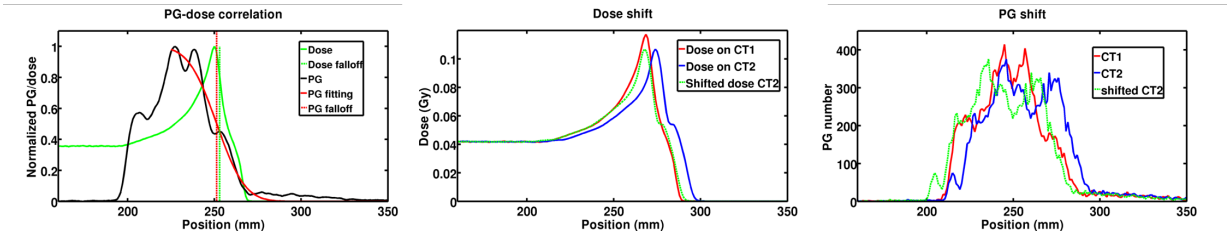


Figure 5.16: The PG-dose correlation quantified from the PG fitting and 80% dose value on CT1, the shift of the dose profiles (-6.2 mm) as well as the shift of the PG profiles (-13.3 mm) on CT1 and CT3 for the metioned PB of figure 5.14 are shown on the left, middle and right, respectively

this study. This dose surface is calculated as follows: 1. The 3D dose distribution of the PB is rotated such that the beam direction is along the left-right (LR) axis; 2. The rotated 3D dose matrix consists of dose voxel columns along the LR axis (figure 5.17 A). The 2D dose resulting from the averaging of the 3D dose in beam direction is used to select dose columns with a value higher than 50% of the maximum averaged dose; 3. The 80% distal falloff position of the LR profile of each dose column is regarded as the range of this dose column (all profiles have the same origin instead of patient surface so that only the geometrical effect of the distal dose distribution, regardless of the penetration depth in tissue, is accounted for); 4. For the 3D dose distribution of a given PB, a 2D surface is composed of the range of each dose column (called dose surface). To further investigate the degree of the tissue heterogeneity in the beam path, the CT numbers (HU) in the voxels within the same dose columns but starting from the patient skin to the 80% distal dose falloff are analyzed in terms of mean and standard deviation (std) values. These CT-based values are not used for the PB identification but only for subsequent analysis.

The effect of tissue heterogeneities on the 3D dose distribution used for the PBs identifi-

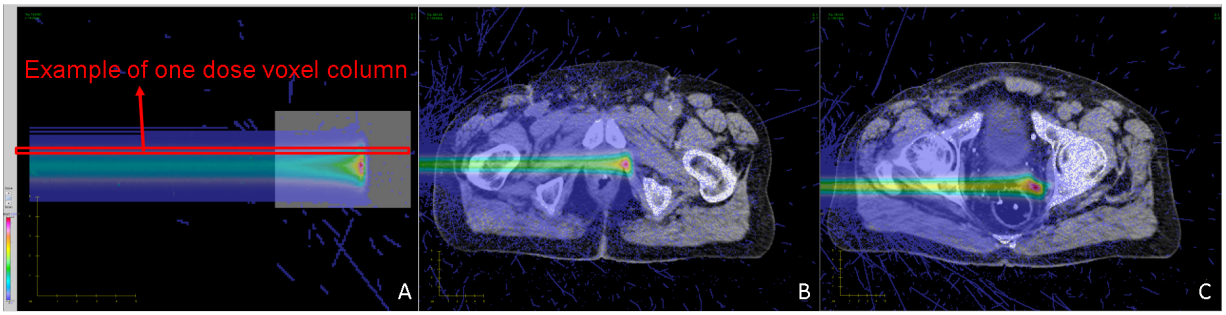


Figure 5.17: 3D MC dose distribution of a given PB passing through (A) water and (B) a relatively homogeneous tissue as well as (C) a pronounced heterogeneous tissue in the pelvic region. The S_{std} calculated values are 0.03, 0.50 and 2.8 mm for A, B and C, respectively. The std values of the CT numbers in the beam path for A, B and C are 0, 199 and 219, respectively. (Tian et al. 2020)

cation is quantified by the standard deviation of the dose surface, S_{std}:

$$S_{std} = \sqrt{\frac{\sum_1^n (range_i - range_{mean})^2}{n-1}} \tag{5.12}$$

where $range_i$ is the range of the dose column i, $range_{mean}$ is the mean value of the ranges of all considered dose columns, and n is the number of the columns under consideration. A small S_{std} value means that the dose columns have similar ranges, i.e. the dose distribution of a given PB is not distorted due to the tissue heterogeneities (see figure 5.17 A and B). Thus, the dose surface could be used as indicator of PBs more sensitive to anatomical changes.

5.5.4 PB selection accounting for the dose surface

The PG-dose correlation and the dose surface indicators are combined to select PG friendly PBs for boosting (figure 5.18). The criteria for the PG-dose correlation are the same as described in section 5.3. For the dose surface indicator, the threshold of the S_{std} value is set to 1.5 mm, slightly higher than the desired PG precision of 1 mm, based on the distribution of the S_{std} values of the PBs investigated. Considering the potential anatomical changes, e.g. the movement of heterogeneities, only PBs with small S_{std} values, and which are surrounded by other PBs with acceptable S_{std}, i.e. PBs whose center does not fall within a 5 mm safety ring of any PB with a S_{std} value larger than 1.5 mm, are suggested for selection. The value of the safety margin of 5 mm is set the same as the spacing between PBs used in the TP.

For each field of the considered patients, 5 PBs recommended by the new approach (called

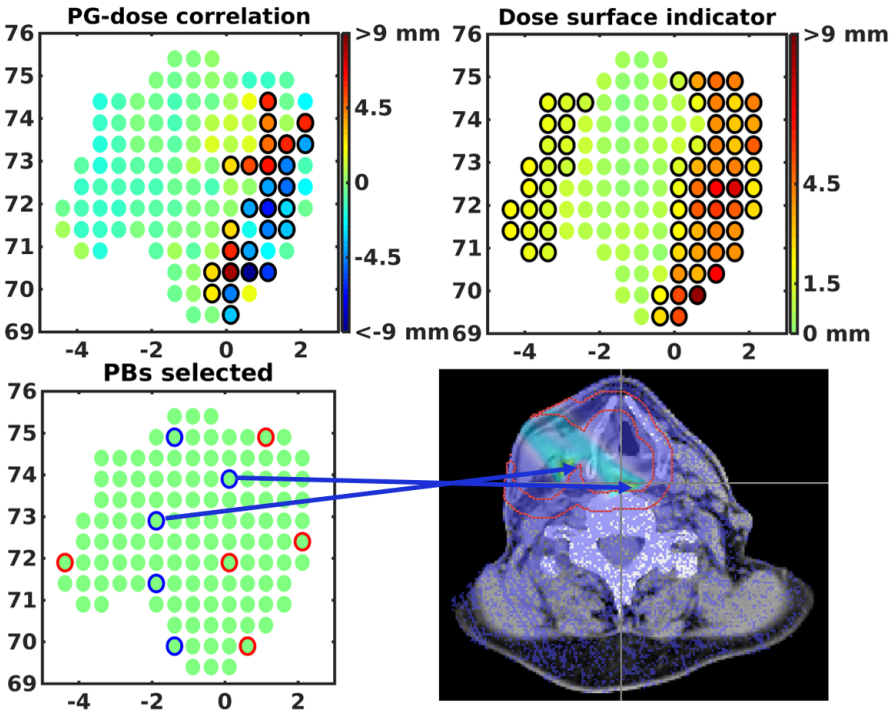


Figure 5.18: The PBs selection combining the PG-dose correlation (top left) and dose surface (top right) indicators. The ΔR_i (top left) and S_{std} (top right) are color coded and the spots without a black circle are PBs recommended by the corresponding indicator. The PBs selected and boosted in this study are shown on the bottom left (blue for good PBs and red for counter-indicated PBs). The MC dose distributions of given PBs are shown on the bottom right. (Tian et al. 2020)

good PBs in the following) and 5 counter-indicated PBs rejected by at least one indicator are selected and boosted in the re-optimized TPs for comparison. The w_{boost} is set to 1.8×10^8 and 2.0×10^8, which are values above the statistical threshold and do not spoil

the TP for the investigated patient cases H&N1 and PROST1, respectively.

5.6 Comparison between the re-optimized TP and initial CERR TP

5.6.1 Total dose distribution

The 3D dose distribution of the initial and re-optimized TPs on all CTs are compared in figure 5.19, 5.20 and 5.21. In this particular case of repeated CTs accounting for anatomical changes, the CTV (which describes the clinical extent of the tumour without additional margins to account for uncertainties) was considered. The total dose distributions of all re-optimized TPs, regardless whether good or counter-indicated PBs are boosted, are observed comparable to the initial CERR TP, as the DVHs of all TPs almost overlap for the same volume on the same CT scan. For the CT3 of the prostate patient, hot spots are found at the same position for all TPs, however the DVHs are still comparable.

This result confirms the previous observations that boosting a few PBs above the statistical threshold does not visibly change the TP in terms of the total dose distribution.

5.6.2 Dose averaged LET distributions

Figure 5.22, 5.23 and 5.24 compare the LET_d distributions of all TPs optimized on CT1 and then recalculated on CT1, CT2 and CT3 of the above mentioned patients, generally yielding similar LET_d value distributions with deviations of the maximum values within approximately 1.6%. This is comparable to the findings of the previous work (section 5.4) that the plan re-optimization boosting a few PBs does not significantly change the dose averaged LET distribution.

5.6.3 Shift of dose and PG on different CTs

The spot-by-spot shifts of the laterally integrated dose profiles on different CTs are compared to the corresponding shifts of the laterally integrated PG profiles to evaluate if a PB is reliable for PG-based proton range verification under given anatomical changes.
The comparison of the shifts as well as the proposed dose surface indicator are shown in the figures 5.25, 5.26 and 5.27. Both PB boosting and spot aggregation (cf. section 5.2) approaches are tested in terms of the dose/PG shift estimation for the selected PBs, where recommended/counter-indicated spots are denoted with blue/red circles, respectively. Detailed data of the boosted PBs (both good and counter-indicated) are listed in tables 5.1- 5.4.

From the tables 5.1- 5.4, the maximum dose/PG shifts of good PBs are observed

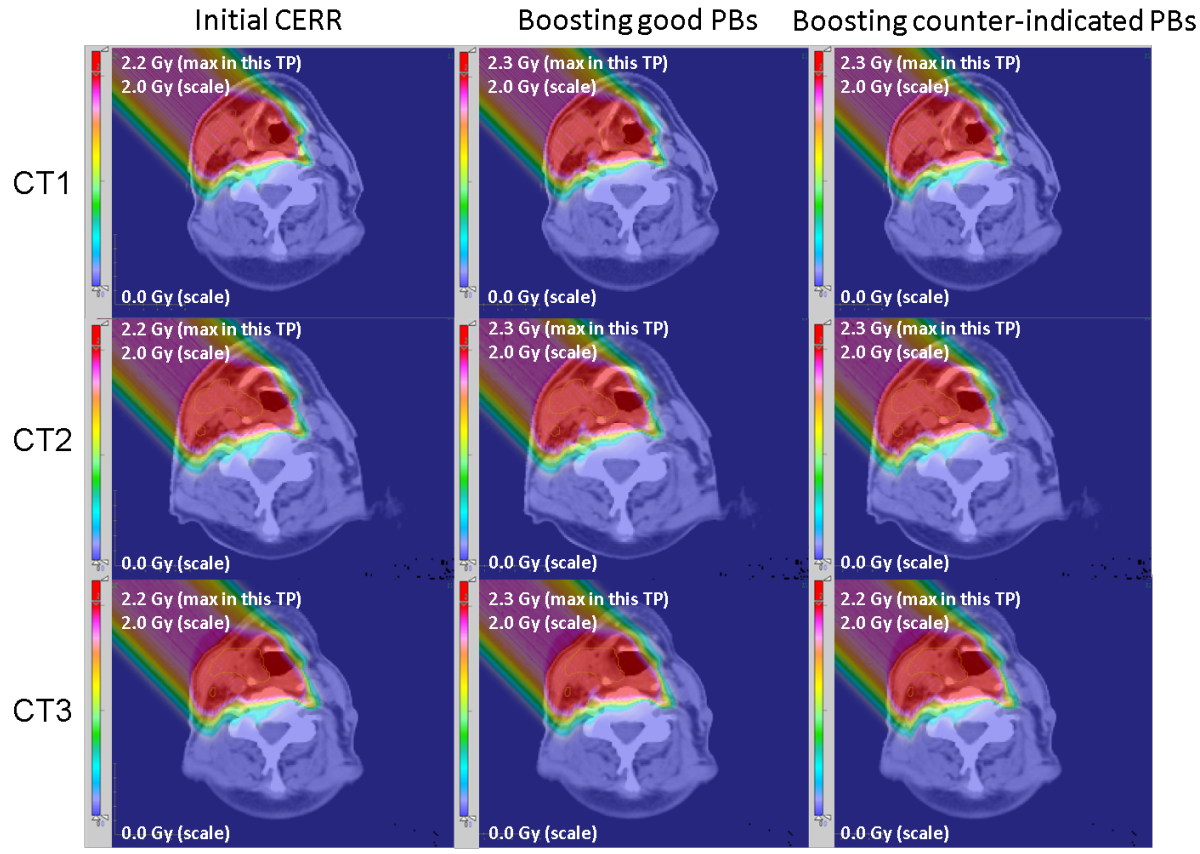


Figure 5.19: Total dose distribution of the initial CERR TP (left), TP boosting good PBs (middle) and TP boosting counter-indicated PBs (right) for the head and neck cancer patient, all optimized on CT1 and then recalculated on the three different CTs acquired at different time points. The scale and maximum dose values are shown next to the color bar. (Tian et al. 2020)

to be 5.2/5.2 mm and 4.2/4.1 mm for the head and neck and prostate cancer patient, respectively. For the good and counter-indicated PBs, the largest difference between the shift of the dose and PG profiles are 0.8 and 5.7 mm, respectively. It can be observed in figure 5.25- 5.27 that all PBs whose PG and dose shift do not match, i.e. the solid red/blue spots in the bottom middle and bottom right BEV, are found to have large S_{std} values (solid red spots in the bottom left) while the PBs recommended by the combined indicators have small shift difference values and are thus more reliable for PG-based proton range verification.

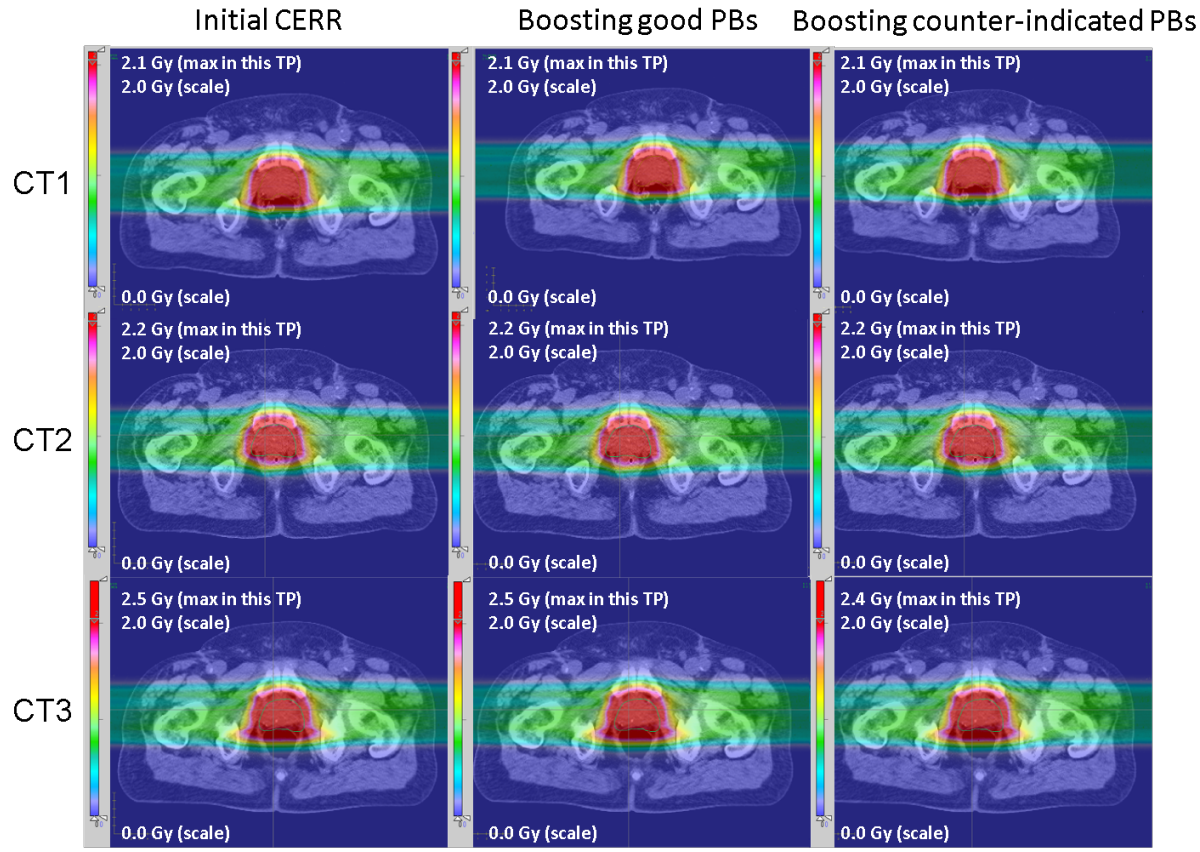


Figure 5.20: Total dose distribution of the initial CERR TP (left), TP boosting good PBs (middle) and TP boosting counter-indicated PBs (right) for the prostate cancer patient, all optimized on CT1 and then recalculated on the three different CTs acquired at different time points. The scale and maximum dose values are shown next to the color bar. (Tian et al. 2020)

5.6.4 Tissue heterogeneity in the beam path

The CT scans with superimposed the dose distributions of two exemplary PBs (one good and one counter-indicated) are shown in figure 5.28. The mean and std values of the CT numbers in the beam path of all the boosted PBs are shown in table 5.5. The std values of the CT numbers in the beam path of the counter-indicated PBs (mean of std: 268) are higher than those of the good PBs (mean of std: 141) for the head and neck patient, while for the prostate patient the CT numbers std values are more comparable for the good (mean of std: 160) and counter-indicated (mean of std: 194) PBs. Hence, the good and counter-indicated PBs do not necessarily correspond to traversed regions of low and high tissue heterogeneity, respectively.

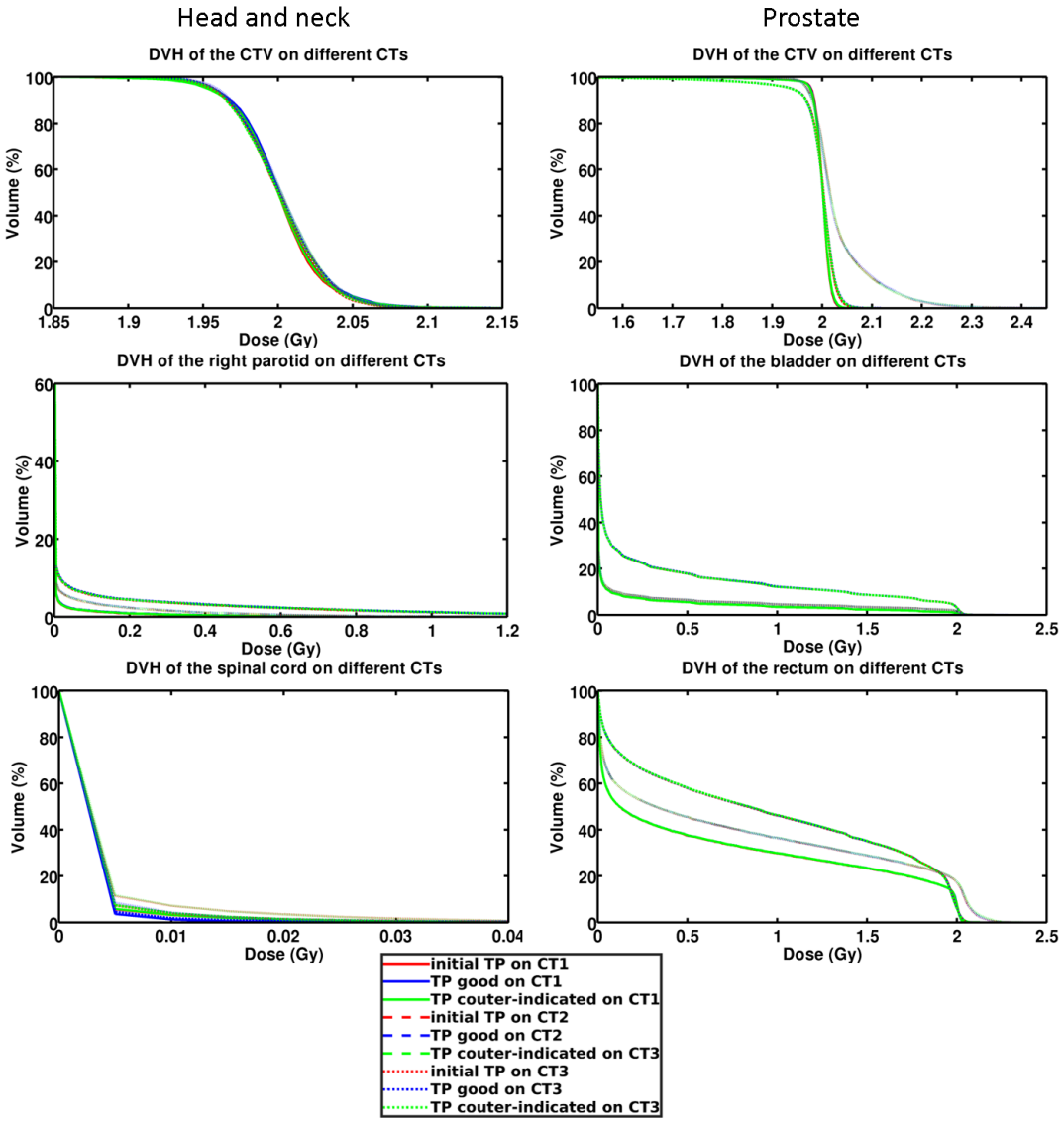


Figure 5.21: The DVHs of the initial CERR TP (red), TP boosting good PBs (blue) and TP boosing counter-indicated PBs (green) devised from the dose recalculated with Monte Carlo on the CT1 used for planning (solid line) as well as on CT2 (dashed line) and CT3 (dotted line) acquired at different time points. The DVHs of the clinical target volume (CTV), and nearby OARs of the head and neck patient are shown on the left while the DVHs of the relevant volumes for the prostate patient are shown on the right. (Tian et al. 2020)

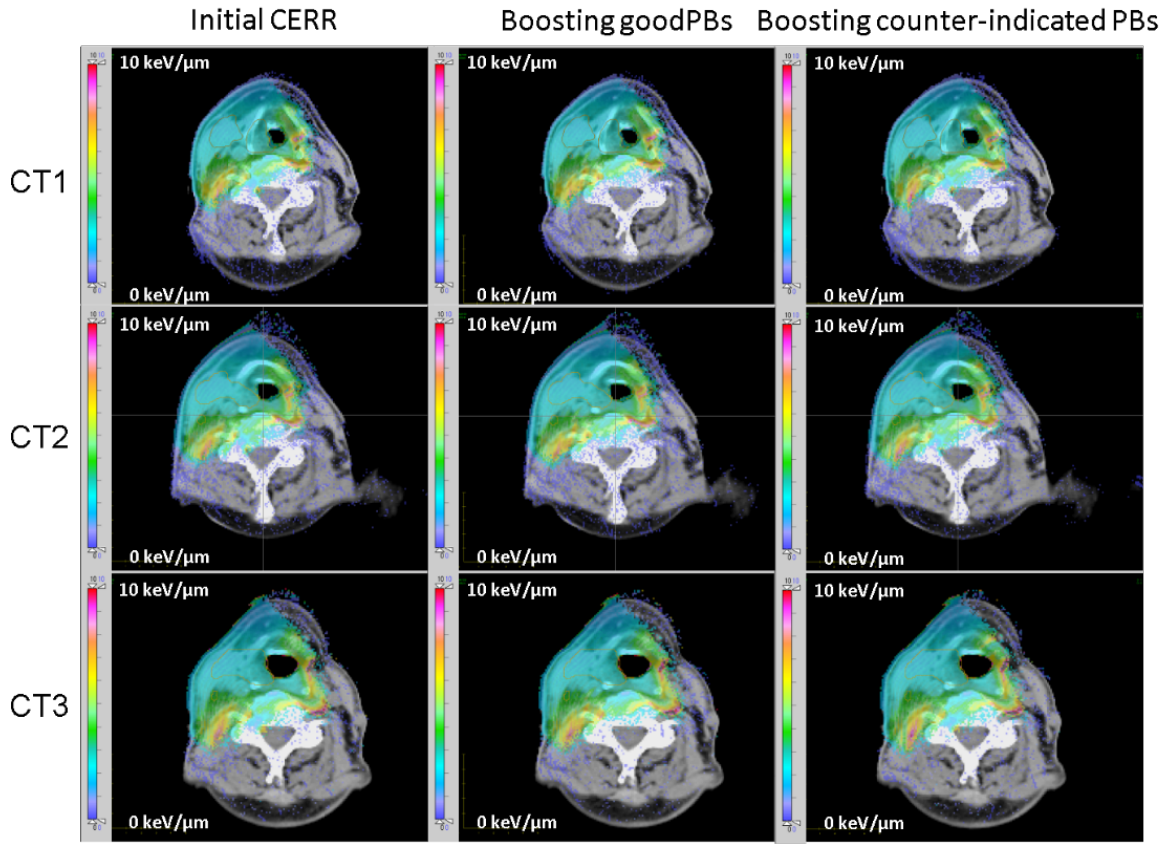


Figure 5.22: The dose averaged LET distribution of the initial TP (left), TP boosting good PBs (middle) and TP boosting counter-indicated PBs (right) for the head and neck cancer patient, all optimized on CT1 and then recalculated on the three different CTs acquired at different time points. (Tian et al. 2020)

5.7 Comparison between the statistics of PB aggregation and PB boosting

In the spot aggregation approach, the aggregated PG profile is compared to the dose of the central PBs, hence it is necessary to check the statistical contribution c of the central PB to the aggregation, calculated by:

$$c = \frac{N_{center}}{N_{aggregation}} \tag{5.13}$$

$$N_{aggregation} = \sum_{i=1}^{m} \phi_i \cdot N_i \tag{5.14}$$

where the N_{center} and $N_{aggregation}$ are the statistics of the central PB and the aggregated PBs, respectively. ϕ_i is the Gaussian weight cf. section 5.2 of PB i, N_i is its statistics in

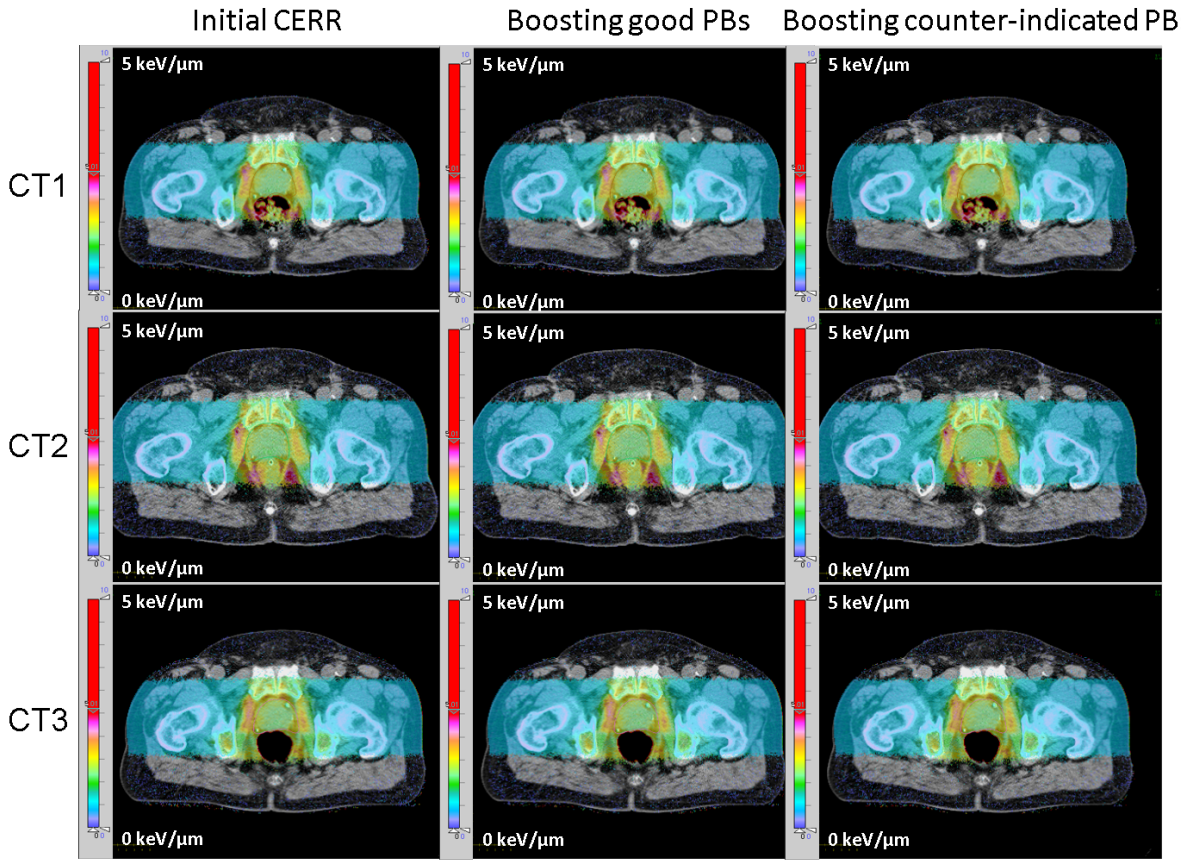


Figure 5.23: The dose averaged LET distribution of the initial TP (left), TP boosting good PBs (middle) and TP boosting counter-indicated PBs (right) for the prostate cancer patient, all optimized on CT1 and then recalculated on the three different CTs acquired at different time points. (Tian et al. 2020)

the initial TP, m is the number of nearby PBs aggregated.
PBs aggregated, with the same incident energy, may have different ranges due to the tissue heterogeneities. To evaluate this effect, the range (all profiles have the same origin) of the dose profile of the central PB is compared to the ranges of each nearby PBs aggregated. The maximum range difference of a given central PB and its nearby PBs is called range mixing (ΔRange) of this PB.

Table 5.6 and 5.7 evaluate the spot aggregation approach in terms of statistics, range mixing and the contribution of the central PB for 15 recommended PBs and 15 counter-indicated PBs from the initial TP for the H&N and the prostate cancer patient. The statistics of the boosted PBs is always higher than the set w_{boost} (1.8×10^8 and 2.0×10^8 protons for the H&N and prostate patient, respectively), while the statistics of the aggregated PB is not guaranteed to exceed this value. The tissue heterogeneities have a considerable effect on the PB aggregation as a maximum range mixing of 14.6 mm is ob-

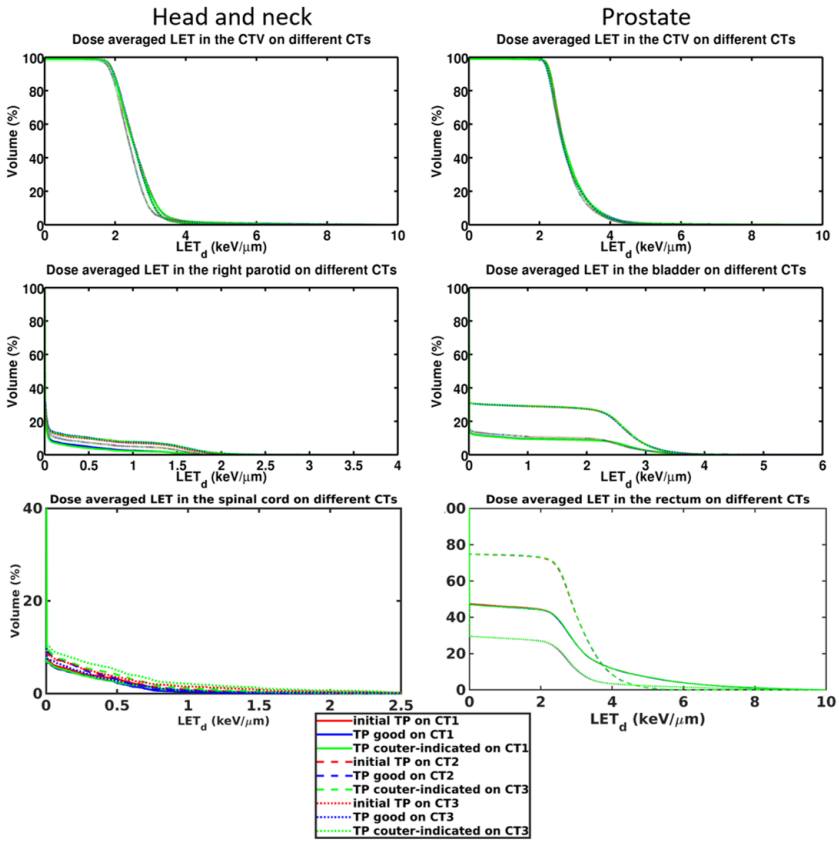


Figure 5.24: The LET_d-volume histogram of the initial TP (red), TP boosting good PBs (blue) and TP boosing counter-indicated PBs (green), all optimized on CT1 and then recalculated on CT1 (solid line), CT2 (dashed line) and CT3 (dotted line). The DVHs of the CTV and nearby OARs of the head and neck patient are shown on the left while the DVHs of the relevant volumes for the prostate patient are shown on the right. (Tian et al. 2020)

served. Furthermore, the statistics of the central PBs in the initial CERR TP is not always high enough to contribute dominantly to the aggregation.

Though the performance of the spot aggregation depends on the specific TP and specific PBs aggregated, it can be inferred from our investigations that using only the spot aggregation may not be sufficiently satisfactory. On the one hand, the statistics of the aggregation might still not be above the PG detectability threshold. On the other hand, the spot aggregation is limited by the effect of tissue heterogeneities, resulting in range mixing, and loss of lateral resolution. Those two issues can be addressed by the PB boosting approach proposed in this thesis work.

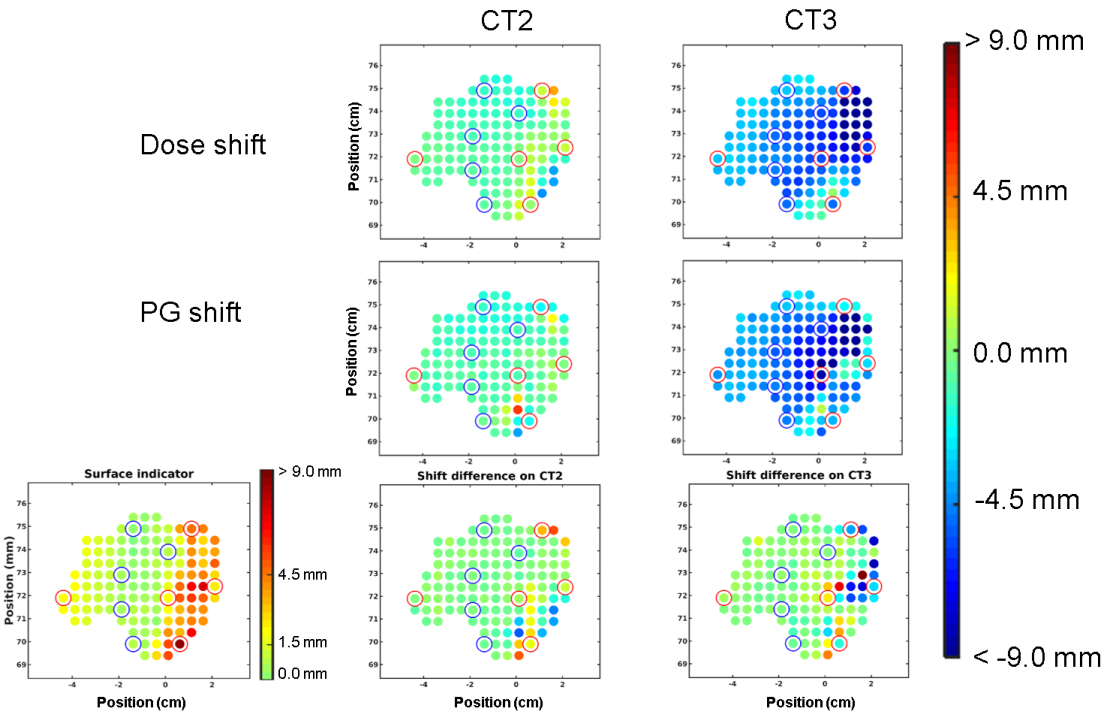


Figure 5.25: BEV of the PBs used for the head and neck cancer patient. The shift comparing the dose on CT2/CT3 to the dose on CT1 is shown in the top row. The middle row shows the corresponding PG shift. The dose surface is shown on the bottom left. The bottom middle and bottom right BEV show the difference between the shift of the dose and PG on the different CTs with respect to the initial CT1 used for planning. The spots wih blue/red circles are the good/counter-indicated PBs boosted in the re-optimized TP. (Tian et al. 2020)

5.8 Discussion

In this part of the work, two indicators are built to quantify the spot-by-spot PG-dose correlation as well as the effect of tissue heterogeneities on the dose distribution. TPs are re-optimized by boosting 5 selected PBs per field. The PG-dose correlation indicator is tested using the MC recalculation data on the same patient CT and the dose surface indicator is estimated from the MC simulation on CTs taken at different time points of the same patient. The comparisons show that both the initial TPs and the re-optimized TPs are comparable in terms of TP robustness as well as dose and dose averaged LET distributions on all CTs of the corresponding patients used in this study. PBs recommended by the proposed indicators are observed to have better PG-dose correlation on both planning CT and CTs at different time points in the course of fractionated radiotherapy, as the

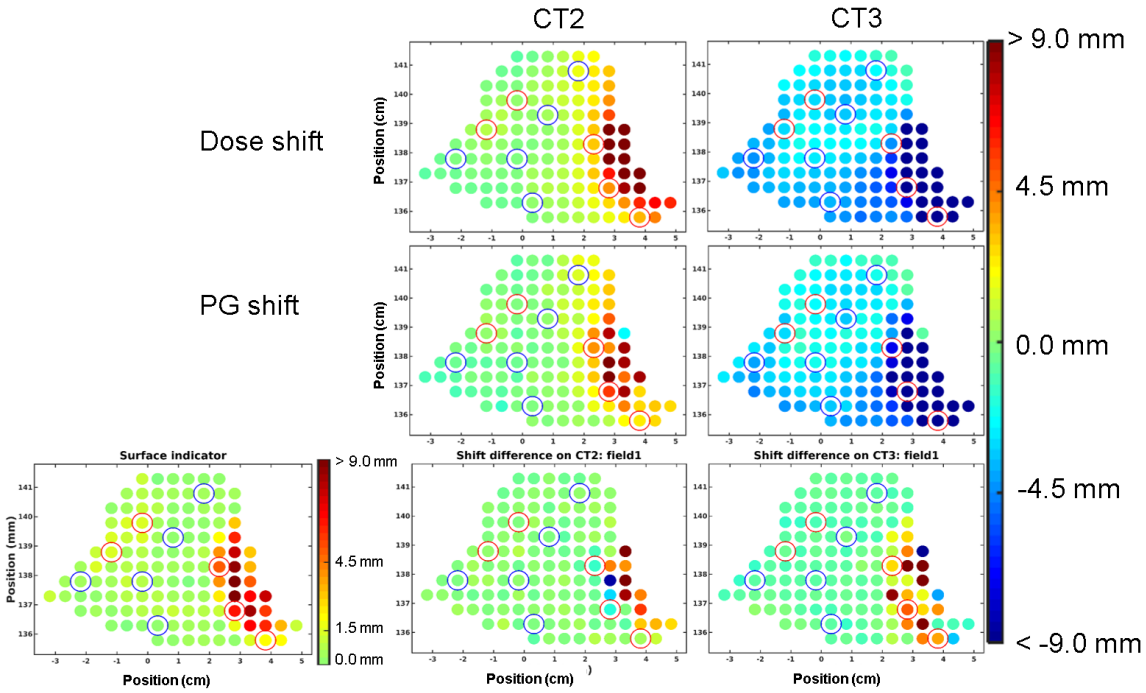


Figure 5.26: BEV of the PBs used in the field 1 for the prostate cancer patient. The shift comparing the dose on CT2/CT3 to the dose on CT1 is shown in the top row. The middle row shows the corresponding PG shift. The dose surface is shown on the bottom left. The bottom middle and bottom right BEV show the difference between the shift of the dose and PG on the different CTs with respect to the initial CT1 used for planning. The spots wih blue/red circles are the good/counter-indicated PBs boosted in the re-optimized TP. (Tian et al. 2020)

largest dose shift observed is 5.2 mm and the maximum difference between the shift of the dose and PG in the planning CT and additional CTs is 0.8 mm. Moreover, PBs boosted are guaranteed to have enough statistics for PG detection. This new planning approach inherently lends itself to the possible future implementation of range verification by first delivering and reliably monitoring the boosted PBs, which are part of the TP, without compromising the plan dosimetric quality. Furthermore, the concept proposed here in the context of PG could also be applied to other technologies than PG imaging.
The boosting planning approach critically relies on the ability to identify those spots which offer the best PG-dose correlation at the time of the initial planning. To this end, a machine learning based pattern recognition algorithm was introduced to augment the traditional fitting method for the falloff identification. The neural network was trained using more than 10000 PBs with good PG-dose correlation. The accuracy of the algorithm was found

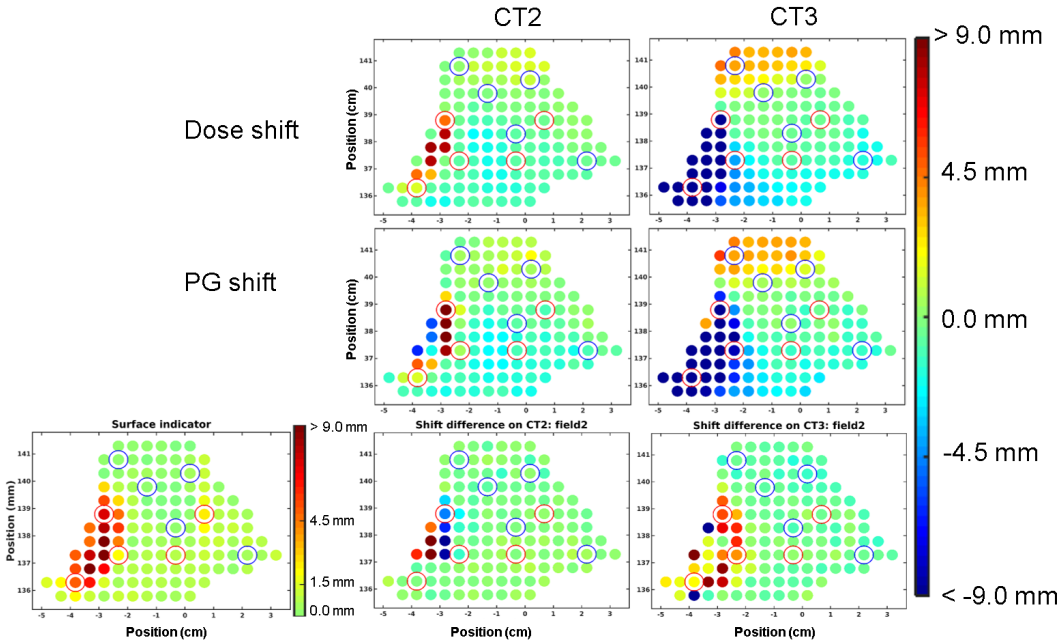


Figure 5.27: BEV of the PBs used in the field 2 for the prostate cancer patient. The shift comparing the dose on CT2/CT3 to the dose on CT1 is shown in the top row. The middle row shows the corresponding PG shift. The dose surface is shown on the bottom left. The bottom middle and bottom right BEV show the difference between the shift of the dose and PG on the different CTs with respect to the initial CT1 used for planning. The spots wih blue/red circles are the good/counter-indicated PBs boosted in the re-optimized TP. (Tian et al. 2020)

higher than 99%, as it only failed in less than 20 cases among the entire tested 28286 PBs, while for the PBs with good PG-dose correlation the accuracy reached 100%. The procedure of merging falloff 1 and falloff 2 (cf. section 5.3.2) was initially employed to identify and correct for potential machine learning errors. It proved to work efficiently when identifying PBs with the most preferred PG-dose correlation, since the coefficients for identifying the PG-dose correlation of a given PB can be calculated by only one $f_1(x)$ fitting and one $f_0(x)$ fitting, while other calculations are avoided.
More extensive computational resources are required for the MC simulation, especially when the PG information is needed, as more protons are required for adequate statistics for the PG emission signal in comparison to the dose. PG emissions were calculated with high statistics (5×10^5 primary protons) for all the PBs of patient H&N1. Those data were used for the construction of the machine learning model. For other patients, only PBs of interest, i.e. PBs with maximum proton energy which have the Bragg peak inside

Patient & PB ID	ΔDose	ΔPG (aggregation)	ΔPG (boost)	Mean ΔPG (aggregation)	Mean ΔPG (boost)
PROST1 field1 1	2.1	1.7	1.7	0.4	0.4
PROST1 field1 2	0.8	0.7	0.4	0.1	0.4
PROST1 field1 3	0.1	-0.5	-0.2	0.6	0.3
PROST1 field1 4	0.1	-0.4	-0.4	0.5	0.5
PROST1 field1 5	0.3	0.3	0.2	0.0	0.1
PROST1 field2 1	1.0	0.6	0.8	0.4	0.2
PROST1 field2 2	0.4	0.5	0.6	-0.1	-0.2
PROST1 field2 3	-0.5	-0.4	-0.4	-0.1	-0.1
PROST1 field2 4	-0.3	-0.6	-0.9	0.3	0.6
PROST1 field2 5	1.4	1.3	1.0	0.1	0.4
H&N4 field1 1	-1.2	-1.1	-1.2	-0.1	0.0
H&N4 field1 2	-1.3	-1	-0.7	-0.3	-0.6
H&N4 field1 3	-0.8	-0.7	-0.9	-0.1	0.1
H&N4 field1 4	-0.5	-0.4	-0.5	-0.1	0.0
H&N4 field1 5	-0.5	-0.5	-0.2	0.0	-0.3

Table 5.1: Comparison of the shift of the dose and PG laterally integrated profiles for the good PBs on CT2 and CT1 for both patient cases. The shift of the dose, the aggregated PG and the PG of the boosted central PB along with the shift difference between the dose and aggregated PG as well as PG of the boosted central PB are given by ΔDose, ΔPG(aggregation), ΔPG(boost), ΔShift(aggregation) and ΔShift(boost), respectively. The units are mm.

the PTV, were simulated with the same high statistics of 5×10^5 primary protons, while other PBs were calculated using the statistics suggested by the CERR MC TPS handbook (Resch et al. 2017), i.e. 1×10^5 protons per PB, since PG emissions of those PBs are not needed and only dose was scored. The machine learning model trained using the data of patient H&N1 was tested on all the other patients (H&N2, H&N3, H&N4 and PROST1), including head and neck as well as prostate cases, and was observed to have similar accuracy. This is because the machine learning model constructed in this work is trained using data generated from comparably heterogeneous head and neck cases. Hence, there is no need to train new models for a comparably homogeneous anatomy such as the prostate.
The study of (Xie et al. 2017), using a knife-edge slit collimator PG camera, reports that the precision of PG signal depends on the proton statistics and a precision of 2 mm can be achieved when the PB has 2×10^8 protons. Other researcher point out that proton range shifts of 3 mm can be retrieved for a PB with 1×10^8 protons, using a 3D single Compton camera (cf. section 3.1.1). In our study, a precision of 1 mm for the PG phsp data profile is achieved when the proton statistics is boosted above at least 1.35×10^8. This is consistent with the reported values from the literature. The setting of the w_{boost}

Patient & PB ID	ΔDose	ΔPG (aggregation)	ΔPG (boost)	Mean ΔPG (aggregation)	Mean ΔPG (boost)
PROST1 field1 1	-2.1	-1.7	-1.6	-0.4	-0.5
PROST1 field1 2	-3.0	-2.3	-3.0	-0.7	0.0
PROST1 field1 3	-4.0	-3.6	-3.5	-0.4	-0.5
PROST1 field1 4	-2.4	-2.4	-1.9	0.0	-0.5
PROST1 field1 5	-3.5	-2.8	-3.0	-0.7	-0.5
PROST1 field2 1	4.2	4.0	4.1	0.2	0.1
PROST1 field2 2	1.0	0.7	1.2	0.3	-0.2
PROST1 field2 3	-0.3	-0.2	0.5	-0.1	-0.8
PROST1 field2 4	-1.5	-0.5	-1.2	-1.0	-0.3
PROST1 field2 5	1.1	2.4	1.8	-1.3	-0.7
H&N4 field1 1	-3.0	-3.0	-3.2	0.0	0.2
H&N4 field1 2	-5.2	-4.8	-5.2	-0.4	0.0
H&N4 field1 3	-4.6	-4.6	-4.9	0.0	0.3
H&N4 field1 4	-4.2	-4.4	-4.3	0.2	0.1
H&N4 field1 5	-4.6	-4.4	-4.3	-0.2	-0.3

Table 5.2: Comparison of the shift of the dose and PG laterally integrated profiles for the good PBs on CT3 and CT1 for both patient cases. The shift of the dose, the aggregated PG and the PG of the boosted central PB along with the shift difference between the dose and aggregated PG as well as PG of the boosted central PB are given by ΔDose, ΔPG(aggregation), ΔPG(boost), ΔShift(aggregation) and ΔShift(boost), respectively. The units are mm.

is not arbitrary and a patient-specific tradeoff should be considered between the PG detectability and TP quality. In this work, the boosting weights w_{boost} were set to 1.35×10^8, 1.35×10^8, 1.35×10^8, 1.80×10^8 and 2.00×10^8 for four head and neck cancer and one prostate patient, respectively. Those values were taken from the highest-statistics PBs in the corresponding initial TPs. For future implementation a good tradeoff will be required above the minimum PG detectability threshold for the available detector solutions and the constraints that the chosen w_{boost} should not spoil the TP optimization, e.g. by causing hot spots.

The PB boosting approach has been compared to the PB aggregation method proposed by (Xie et al. 2017). In the study on the influence of anatomical changes, both approaches successfully predicted the dose shift using the PG shift for the PBs recommended by the combined PG-dose correlation and dose surface indicators, while for the counter-indicated PBs both approaches failed. However, the PBs recommended by the indicators introduced in this work and boosted above the statistical threshold are more likely to provide reliable PG information for proton range monitoring. Monitoring an arbitrary PB may not provide correct proton range information. Moreover, the aggregated statistics is not always above

Patient & PB ID	ΔDose	ΔPG (aggregation)	ΔPG (boost)	Mean ΔPG (aggregation)	Mean ΔPG (boost)
PROST1 field1 1	0.5	0.2	0.3	0.3	0.2
PROST1 field1 2	1.0	0.4	0.4	0.6	0.6
PROST1 field1 3	3.2	3.6	4.3	-0.4	-1.1
PROST1 field1 4	4.4	5.1	5.8	-0.7	-1.4
PROST1 field1 5	3.6	3.0	2.7	0.6	0.9
PROST1 field2 1	-0.2	-0.2	-0.6	0.0	0.4
PROST1 field2 2	5.1	2.5	9.1	2.6	-4.0
PROST1 field2 3	-0.8	1.4	0.6	-2.2	-1.4
PROST1 field2 4	-0.8	-1.4	-0.8	0.6	0.0
PROST1 field2 5	2.1	0.5	1.7	1.6	0.4
H&N4 field1 1	0.3	-1.2	-1.6	1.5	1.9
H&N4 field1 2	1.6	-2.6	-2.1	4.2	3.7
H&N4 field1 3	0.0	-1.0	-0.9	1.0	0.9
H&N4 field1 4	0.6	0.2	0.0	0.4	0.6
H&N4 field1 5	1.4	0.0	0.1	1.4	1.3

Table 5.3: Comparison of the shift of the dose and PG laterally integrated profiles for the counter-indicated PBs on CT2 and CT1 for both patient cases. The shift of the dose, the aggregated PG and the PG of the boosted central PB along with the shift difference between the dose and aggregated PG as well as PG of the boosted central PB are given by ΔDose, ΔPG(aggregation), ΔPG(boost), ΔShift(aggregation) and ΔShift(boost), respectively. The units are mm.

the PG detectability threshold while the statistics of the PBs is intrinsically guaranteed by the PB boosting method. Furthermore, the central PB is not guaranteed to dominantly contribute to the aggregation without special TP re-optimization. For instance, the third counter-indicated PB used for one of the head and neck cancer patient, cf. table 5.3 and 5.4, has only 2.4×10^6 protons while one of its aggregated nearby PB (more than 10 mm away, with a Gaussian weight about 0.36) has 3.0×10^7 protons in the initial TP. In this scenario the aggregated PG signal may reflect information more about that nearby PB instead of the central one. Besides, the nearby PBs aggregated may have different proton range due to the tissue heterogeneities (a distal surface difference of 14 mm has been observed). These effects are not taken into account in the PBs aggregation approach. The approach investigated in this thesis shows advantages in terms of guaranteed statistics, reduced range mixing, improved lateral resolution and the contribution of the central PB. In any case there is no conflict between those two approaches, as PG data of individual PBs could still be aggregated regardless whether the central PBs are boosted or not. In contrast, using the proposed approach to boost a few selected PBs could reduce the effects mentioned above if the two approaches are combined.

Note that the dose surface indicator is introduced to quantify the effect of tissue hetero-

Patient & PB ID	ΔDose	ΔPG (aggregation)	ΔPG (boost)	Mean ΔPG (aggregation)	Mean ΔPG (boost)
PROST1 field1 1	-2.8	-2.3	-1.9	-0.5	-0.9
PROST1 field1 2	-2.7	-1.9	-2.7	-0.8	0.0
PROST1 field1 3	-3.9	-4.7	-6.9	0.8	3.0
PROST1 field1 4	-15.6	-20.8	-20.8	5.2	5.2
PROST1 field1 5	-19.3	-25.1	-23.4	5.8	4.1
PROST1 field2 1	-0.4	-0.1	-0.2	-0.3	-0.2
PROST1 field2 2	-8.6	-3.9	-14.3	-4.7	5.7
PROST1 field2 3	-3.8	-10.4	-8.0	6.6	4.2
PROST1 field2 4	-0.8	-1.1	-1.0	0.3	0.2
PROST1 field2 5	-27.2	-26.5	-29.4	-0.7	2.2
H&N4 field1 1	-4.6	-3.5	-3.3	-1.1	-1.3
H&N4 field1 2	-5.8	-2.1	-2.0	-3.7	-3.8
H&N4 field1 3	-5.3	-7.2	-8.1	1.9	2.8
H&N4 field1 4	-2.6	-4.8	-3.9	2.2	1.3
H&N4 field1 5	-5.6	-3.5	-2.8	-2.1	-2.8

Table 5.4: Comparison of the shift of the dose and PG laterally integrated profiles for the counter-indicated PBs on CT3 and CT1 for both patient cases. The shift of the dose, the aggregated PG and the PG of the boosted central PB along with the shift difference between the dose and aggregated PG as well as PG of the boosted central PB are given by ΔDose, ΔPG(aggregation), ΔPG(boost), ΔShift(aggregation) and ΔShift(boost), respectively. The units are mm.

geneities and is mainly sensitive to the transversal tissue heterogeneities. It does not always reject PBs delivered in heterogeneities. A large part of range uncertainties is caused by systematic errors, e.g. in the CT-range calibration or by gross body surface and anatomical changes. For those uncertainties, only a few PBs with good PG-dose correlation and small dose surface std should be enough to give sufficient information whether the actual irradiation scenario is consistent with the planning situation. For the PBs rejected by the indicators, the shift of the PG is not guaranteed to match the shift of the dose accurately. However, for most of the investigated PBs a large PG shift was typically observed when the shift of the dose was large, and the difference between the shift of the dose and PG was small. Hence, even for counter-indicated PBs, although the dose shift cannot be inferred perfectly by the detected PG shift, PG monitoring could still be valuable to trigger additional investigations, e.g. a new CT. The proposed indicators only provide the information whether the dose shift of a given PB can be predicted accurately from the PG shift or not. The user can choose and boost the PBs of interest, e.g. the good ones to monitor more systematic range errors in regions of reduced transversal heterogeneities, or the counter-indicated ones which are expected to be more sensitive to range uncertainties,

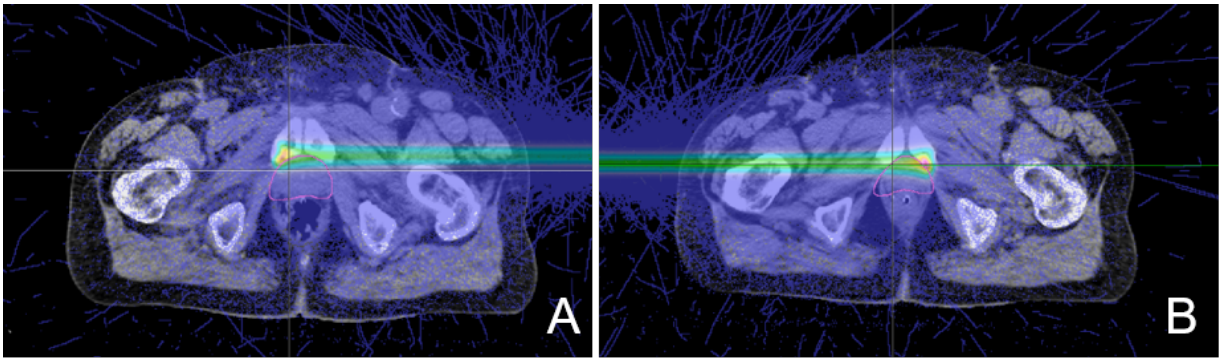


Figure 5.28: The CT scans with superimposed MC dose distributions of a good and a counter-indicated PB are shown on the left and right, respectively. The std values of the CT numbers for A and B are 96 and 236, respectively. (Tian et al. 2020)

while maintaining the TP quality.
A safety ring of 5 mm was applied for the identification of the good PBs, i.e. only PBs surrounded by small-dose-surface-std PBs were recommended, to prevent loss of robustness of the PG-dose correlation from slight anatomical changes. However, the anatomical changes are hard to predict. Though all PBs recommended in this work show small difference of the shift between the dose and PG while the shift of the dose can reach 5.2 mm, the indicators might still fail in the case of dramatic anatomical changes.
The PG signal for the work described in this chapter was obtained using the ideal emission and phsp data calculated by Geant4, neglecting the response of a PG detection system (cf. different solutions reviewed in chapter 3). A relevant study to address this aspect is presented in chapter 6.

Patient & PB ID (good)	Heterogeneities (HU)	PB id (counter-indicated)	Heterogeneities (HU)
PROST1 field1 1	50 ± 192	Prostate field1 1	38 ± 177
PROST1 field1 2	40 ± 186	Prostate field1 2	18 ± 96
PROST1 field1 3	29 ± 123	Prostate field1 3	48 ± 243
PROST1 field1 4	43 ± 137	Prostate field1 4	-3 ± 225
PROST1 field1 5	47 ± 130	Prostate field1 5	17 ± 197
PROST1 field2 1	25 ± 134	Prostate field2 1	29 ± 154
PROST1 field2 2	39 ± 181	Prostate field2 2	-26 ± 291
PROST1 field2 3	43 ± 151	Prostate field2 3	50 ± 199
PROST1 field2 4	23 ± 129	Prostate field2 4	40 ± 134
PROST1 field2 5	71 ± 237	Prostate field2 5	-26 ± 219
H&N4 field1 1	5 ± 140	H&N 1	-185 ± 432
H&N4 field1 2	28 ± 156	HH&N 2	-123 ± 326
H&N4 field1 3	0 ± 136	H&N 3	-26 ± 204
H&N4 field1 4	-34 ± 147	H&N 4	-47 ± 169
H&N4 field1 5	-32 ± 126	H&N 5	-40 ± 211

Table 5.5: Tissue heterogeneities in the beam path of the boosted good and counter-indicated PBs (mean std value)

Patient & PB ID	ΔRange (mm)	$N_{\text{aggregation}}$	N_{central}	C
PROST1 field1 1	2.9	3.1×10^8	7.4×10^7	23.8%
PROST1 field1 2	2.6	1.9×10^8	1.4×10^7	7.3%
PROST1 field1 3	3.0	1.7×10^8	2.8×10^7	16.4%
PROST1 field1 4	3.8	4.0×10^8	4.7×10^7	11.8%
PROST1 field1 5	4.0	1.4×10^8	2.4×10^6	1.7%
PROST1 field2 1	1.4	1.9×10^8	4.6×10^7	24.2%
PROST1 field2 2	1.8	7.7×10^7	4.0×10^6	5.2%
PROST1 field2 3	4.3	2.6×10^8	4.1×10^7	15.8%
PROST1 field2 4	4.2	6.2×10^7	8.4×10^5	1.4%
PROST1 field2 5	5.3	1.3×10^8	7.8×10^6	6.0%
H&N4 field1 1	3.6	2.1×10^8	7.7×10^5	0.4%
H&N4 field1 2	4.1	2.4×10^8	2.1×10^7	8.9%
H&N4 field1 3	4.6	1.4×10^8	8.2×10^6	5.9%
H&N4 field1 4	3.9	1.3×10^8	1.1×10^6	0.9%
H&N4 field1 5	3.9	1.5×10^8	4.1×10^5	0.3%

Table 5.6: The range mixing and statistics of the PB aggregation for the selected good PBs are shown. The range mixing, aggregated statistics, statistics of the central PB in the initial TP and the corresponding contribution are given by ΔRange, $N_{\text{aggregation}}$, N_{central}, and C, respectively.

Patient & PB ID	ΔRange (mm)	$N_{\text{aggregation}}$	N_{central}	C
PROST1 field1 1	3.7	2.2×10^8	3.2×10^7	14.4%
PROST1 field1 2	3.0	4.0×10^7	2.7×10^6	6.6%
PROST1 field1 3	4.7	1.9×10^8	4.4×10^7	23.1%
PROST1 field1 4	5.0	1.2×10^8	9.1×10^7	76.2%
PROST1 field1 5	3.8	6.6×10^7	5.1×10^7	76.7%
PROST1 field2 1	3.4	1.5×10^8	5.1×10^6	3.4%
PROST1 field2 2	12.5	5.8×10^8	9.1×10^6	15.7%
PROST1 field2 3	1.7	9.1×10^7	4.5×10^6	5.0%
PROST1 field2 4	4.5	2.0×10^8	4.7×10^7	23.6%
PROST1 field2 5	7.0	1.4×10^8	3.2×10^6	2.3%
H&N4 field1 1	14.6	1.8×10^8	4.1×10^5	0.2%
H&N4 field1 2	10.8	1.2×10^8	3.1×10^5	0.3%
H&N4 field1 3	9.3	2.3×10^8	2.4×10^6	1.0%
H&N4 field1 4	7.8	2.3×10^8	1.4×10^6	0.6%
H&N4 field1 5	9.2	1.4×10^8	6.8×10^5	0.5%

Table 5.7: The range mixing and statistics of the PB aggregation for the selected counter-indicated PBs are shown. The range mixing, aggregated statistics, statistics of the central PB in the initial TP and the corresponding contribution are given by ΔRange, $N_{\text{aggregation}}$, N_{central}, and C, respectively.

Chapter 6

TP re-optimization and PG monitoring considering the response of a knife-edge slit camera prototype

The work reported in the previous chapter focused on the correlation between the laterally integrated depth profiles of dose and PG emission/phsp data simulated in Geant4, not accounting for the response of a real PG camera. On the other hand, the computational platform REGGUI (Janssens et al. 2009) provides the possibility to simulate the response of the knife-edge PG camera prototype used in the first clinical PG application for PB scanning conducted by (Xie et al. 2017). The relevant processes for simulation of dose, PG emission and camera response in REGGUI were described in chapter 3. Since reproduction of such a detector response in the previously used Geant4 computational platform would require access to confidential proprietary camera information, in this study REGGUI was used to investigate how the PG-dose correlation at emission (chapter 5) holds at detection in presence of anatomical changes, taking into account a realistic PG system. The re-optimization of the TP by boosting a few selected PBs above the set statistical threshold while not spoiling the TP is discussed in chapter 5. This chapter focuses only on the performance of the PB selection methods accounting for the camera response. The statistics used is based on either the initial CERR TP (since REGGUI does not support the function of treatment planning) or the preferred statistics of 2×10^8 protons proposed for PG monitoring by the previous study with this camera prototype (Xie et al. 2017).
Moreover, the PG signal detected by this type of system of limited field of view (FOV) is affected by the relative position of the camera and the PG falloff. Hence, in this work, an additional indicator is introduced based on the relative position of the detected PG profile in the field of view (FOV) of the camera, as proposed by (Huang 2020). The PBs used for the TPs of the same patients analyzed in section 5.4-5.5 are thus re-evaluated using the PG-dose correlation, dose surface and camera position indicators, all based on new REGGUI calculations. The combination of the PG-dose correlation and dose surface indicators of chapter 5 is called PG emission indicator in the following. The shifts of the laterally integrated depth profiles of dose, PG emission and PG detection generated

by REGGUI simulation are compared to evaluate if the PBs, selected at emission, still provide correct information for PG-based proton range verification when detected by a real system. Although for each considered patient CTs at three different time points were available, only the planning CT and the additional CT with larger anatomical changes are used (called CT1 and CT2 in the following). Since the effect of boosting a few PBs on the TP has been discussed in the previous investigations and would not change in the newly proposed workflow, this chapter focuses only the reliability of the selected PBs in terms of PG based proton range verification when including the detected signal. PB boosting and PB aggregation are compared.

6.1 Comparison between the dose and PG emission generated by Geant4 and REGGUI

The MC dose engine and the PG emission calculations used by REGGUI are based on the MCsquare code and on the PG database generated by the MC PENH extension to PENELOPE, respectively, which are both different from the Geant4 code (see chapter 4) used in the previous chapter. Hence, the difference between the dose and PG emission calculations of REGGUI and Geant4 has been investigated.

The 3D dose distribution as well as the laterally integrated depth dose and PG emission profiles calculated by Geant4 and REGGUI are compared in figure 6.1 for a given PB delivered near an air cavity in the pelvic region. The shift method as introduced in section 5.5 is employed to quantify the range difference of the profiles calculated by Geant4 and REGGUI. The resulting range differences between profiles calculated by Geant4 and REGGUI are -4.0 mm and -8.6 mm for the dose and PG emission, respectively. The histograms of the dose range difference for all the PBs of interest for the head and neck and prostate patient cases are shown in figure 6.2. A systematic range difference is observed for patient PROST1, due to a comparably longer proton range in Geant4. The PB identification by the PG emission indicators using the different datasets is compared in figure 6.3. The general area of good/counter-indicated PBs are similar. Differences of PBs identifications are mainly observed near the boundary of the recommended and rejected areas. To maintain the data consistency for the investigations presented in this chapter, besides the parameters used for PBs selection that are the same as those used in chapter 5 (based on Geant4), all the data used in the following are generated using REGGUI.

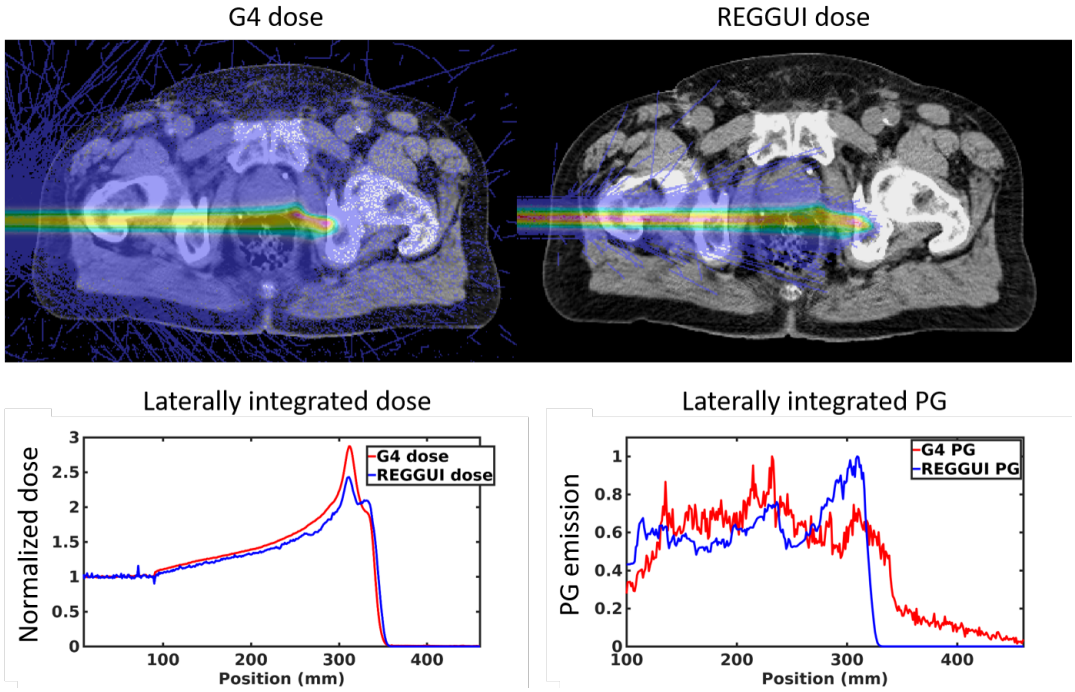


Figure 6.1: The axial slice of the 3D dose distribution in colourwash superimposed onto the grayscale CT (top) as well as laterally integrated depth dose (bottom left) and PG (bottom right) profiles of a PB passing through tissue heterogeneities, as calculated by Geant4 (top: left panel; bottom: red lines) and REGGUI (top: right panel; bottom: blue lines).

6.2 Accounting for camera position

6.2.1 Camera position indicator

Highest-energy candidate PBs with the Bragg peak in the safe region inside the PTV (cf. section 5.3.3) are simulated on CT1 using REGGUI. By giving the camera position with respect to the PTV isocenter, the PG detection profiles consisting of 20 data points, which correspond to the PG counting of 20 detectors (each of 4 mm length) of the camera, are obtained. The profiles are firstly normalized in the region of 17-63 mm (where 0 is the beginning of the first detector) and then smoothed with a Savitzky-Golay filter. The PG profile falloff is defined by the distal region from 90% to 10% of the maximum value. PBs with falloffs which fulfill the following criteria, which were defined based on the analysis of a large number of simulated and measured data by (Huang 2020), are identified to be centered at the camera FOV (see example in figure 6.4):

1. The length of the falloff is longer than 8 mm
2. The falloff is within the region of 18 to 62 mm

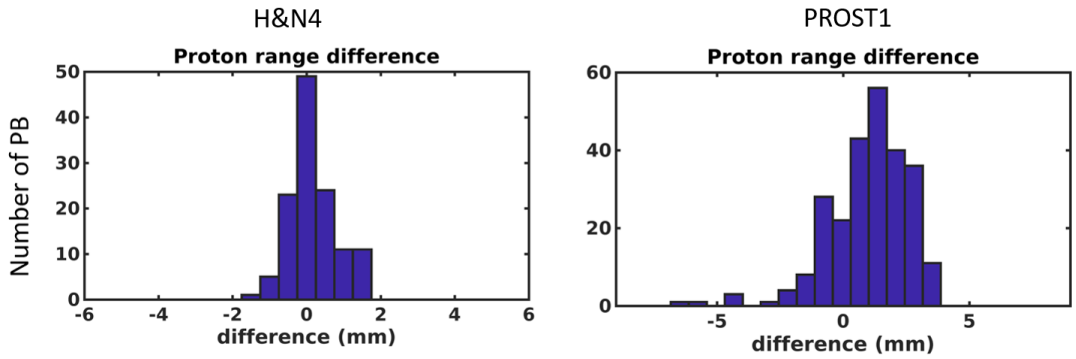


Figure 6.2: Histograms of the proton range differences between the individual PB dose distributions calculated by Geant4 and REGGUI on CT1 for the patients considered in this chapter.

3. The center (50% point) of the falloff is within 36 to 44 mm

Candidate PBs have generally a different range depending on the shape of the PTV, thus not all PBs are necessarily centered in the FOV of the camera at a given position, depending on the position of the camera set by the user. In this work, the camera positions are set to those values where the number of centered PBs is maximized, i.e. 6.25, 9.5 and 48.5 mm away from the PTV isocenter for the TP field of patient H&N4, as well as field 1 and field 2 of patient PROST1, respectively.

6.2.2 Shift calculation

The spot-by-spot shift of the dose, PG emission and PG detection laterally integrated depth profiles on different CTs are compared to evaluate the reliabilities of the selected PBs in terms of proton range verification. The shifts of the dose and PG emission and detection profiles are calculated by applying the same shift algorithm described in chapter 5 on the REGGUI-generated data on the corresponding CTs.
The PG detection profile calculated by REGGUI is the expectation of the PG counting numbers per incident proton in the corresponding camera detectors (called PG simulation in the following), not accounting for the underlying Poisson statistics. To obtain the realistic camera response, the PG simulation profile is firstly multiplied by the PB statistics and converted to integers randomly, e.g. a PG counting of 3.6 is converted to 3 by 40% chance or 4 by 60% chance. This is in the following referred to as PG simulation. A Poisson noise is then added to the counting numbers, thus the camera in-silico measurement profile, hereafter referred to as measurement, is obtained (figure 6.5). In case of emulation

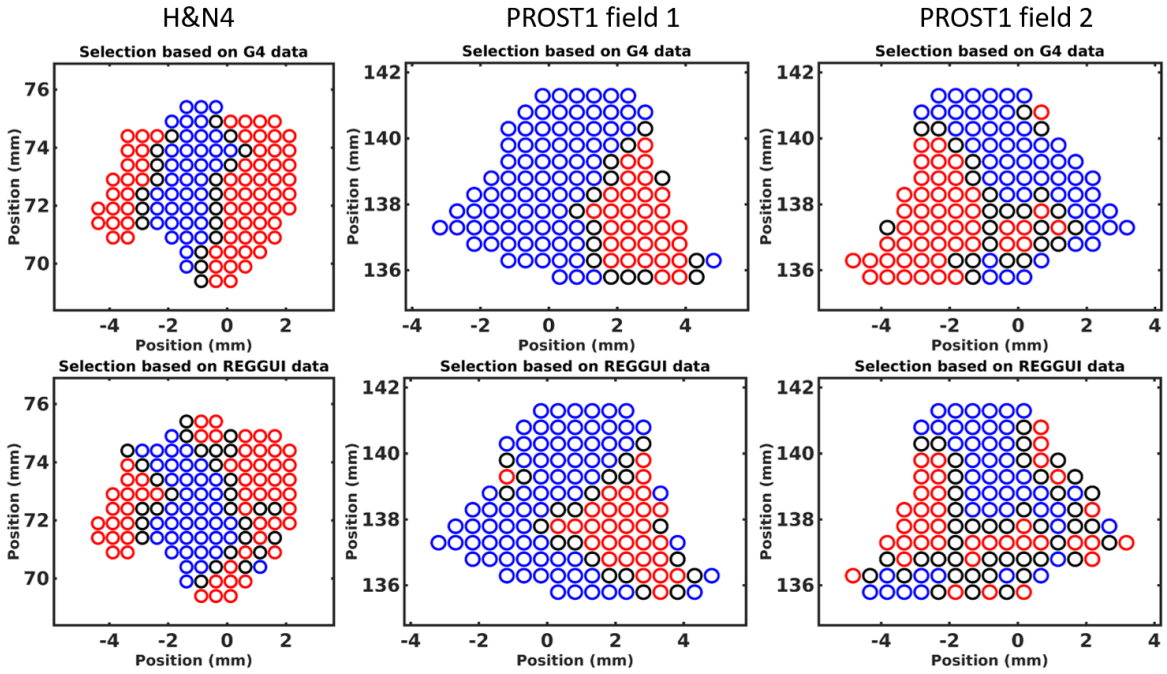


Figure 6.3: PB classification based on the PG emission indicator using Geant4 (top) and REGGUI (bottom) data. The blue and red spots are directly recommended and rejected by the indicator, respectively. The black spots are initially recommended but then rejected when accounting for the safety margin of 5 mm for the dose surface indicator (see section 5.5.4).

of measurements, the aggregation profiles are created by summing up the noisy profiles of relevant PBs using a Gaussian weight, as explained in section 5.2.

The following scenarios are evaluated:

1. Camera response ideal scenario: In this scenario, the shifts difference between dose and PG simulation profiles on different CTs are compared to the shift difference between dose and PG emission to investigate the effect of camera response regardless of statistical noise considerations.
2. Camera response realistic scenario for the comparison between measurement and simulation: 1000 generated PG measurement profiles with Poisson noise on CT2 are compared to the ideal PG simulation profiles on CT1, corresponding to a realistic monitoring workflow which compares the detected PG signal on the daily anatomy to the expected PG simulation on the anatomy from the treatment plan for proton range verification (accuracy check).
3. Camera response realistic scenario for the comparison between measurement

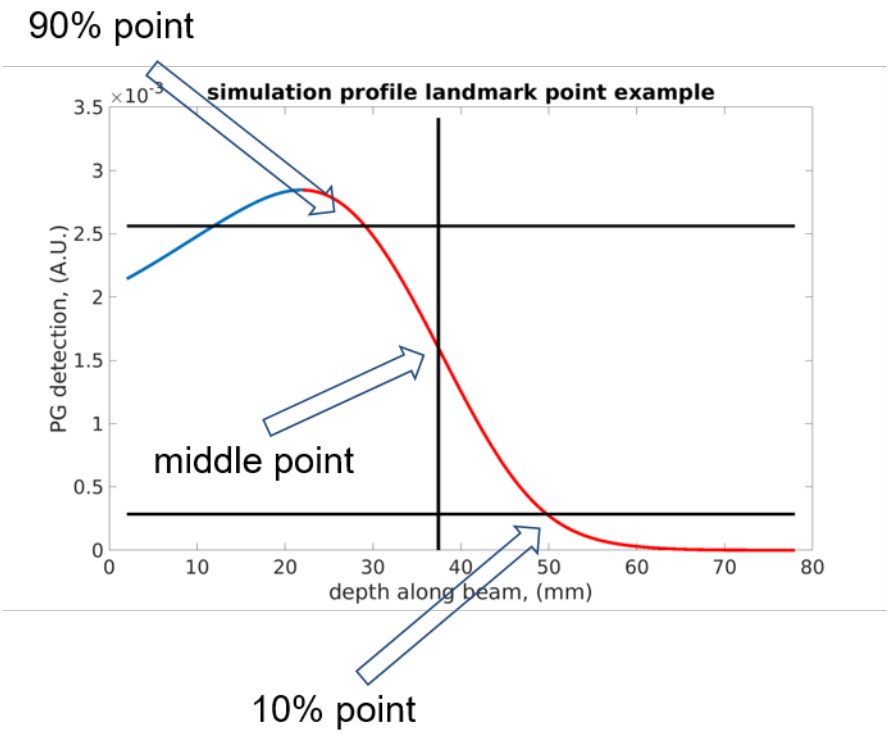


Figure 6.4: The camera position indicator is applied on an exemplary PG detection profile (blue and red). The red line is the falloff region. The 90%, middle (50%) and 10% points are marked by arrows.

and measurement: Shifts of 1000 pairs of PG measurement profiles on different CTs are compared to the dose shifts, corresponding to a realistic monitoring workflow, which compares the detected PG signals at different treatment time points for proton range verification (reproducibility check).

To explore the effects of the counting statistics depending on the amount of primary protons per PB, two scenarios are used:

1. Statistics from the initial CERR TP.
2. Ideal statistics of 2.0×10^8 protons for each PB as suggested by the work of (Xie et al. 2017), to evaluate the implications of the PB boosting method.

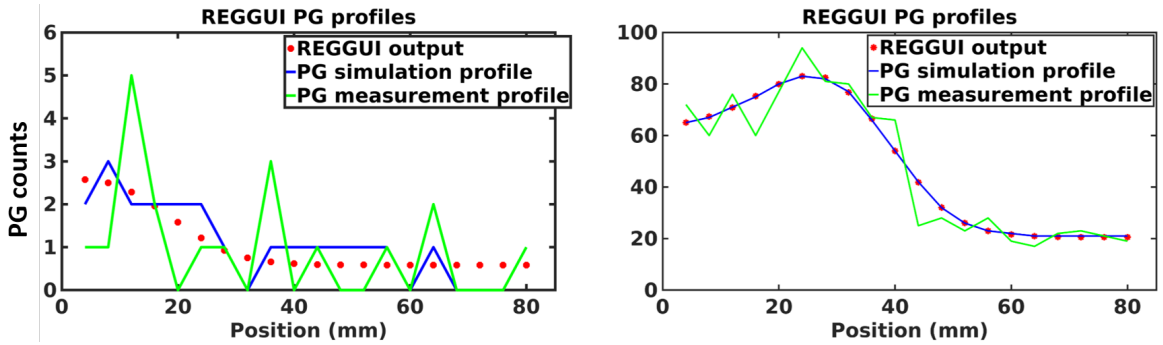


Figure 6.5: The REGGUI output (readout of the 20 detectors of the camera before random rounding), PG simulation profile with random rounding and PG measurement profile accounting for the Poisson noise. The statistics of the left and right profiles are 2.5×10^6 and 6.8×10^7, respectively. The PB statistics used is from the initial TP.

6.3 Results

6.3.1 Effect of PB statistics on the shift detection precision

The maximum PG counts per detector bin generated by the REGGUI simulation are less than 2.1×10^{-6} and 1.7×10^{-6} PG/protons for the patient H&N4 and PROST1, respectively. PBs need more than 5.9×10^5 protons to generate at least 1 PG counting in the camera. A profile with low statistics, e.g. around 5.9×10^5, would consist of a series of small numbers, e.g. 0, 1, 2, 3 etc., and a falloff cannot be identified (see figure 6.5 left, green profile). Comparing 1000 generated PG measurement profiles and the corresponding PG simulation profile, the change of the std value of the shifts following the change of statistics is shown in figure 6.6. A precision of 1.2 mm (1 std) can be achieved when the PB has more than 1×10^8 protons.

6.3.2 Camera response ideal scenario (simulation vs simulation)

In this scenario, the effect of statistical noise is not considered since all signals in this section are deterministic besides the explained random rounding process. The spot-by-spot shifts of the profiles of the dose, PG emission and PG simulation on CT2 with respect to CT1 are compared in figure 6.7 in BEV. Here, colour coding reflects the shift in mm and circles distinguish the identification of the indicators for spot selection: spots with black circles are those recommended by both emission and camera indicators; spots with red/blue circles are those recommended by only the camera/PG emission indicator, respectively; spots without circles are those rejected by both indicators.

The shift differences between the dose and PG emission of most (118/125) PBs recommended by the PG emission indicator are lower than 1.5 mm with a mean value of 0.6 mm

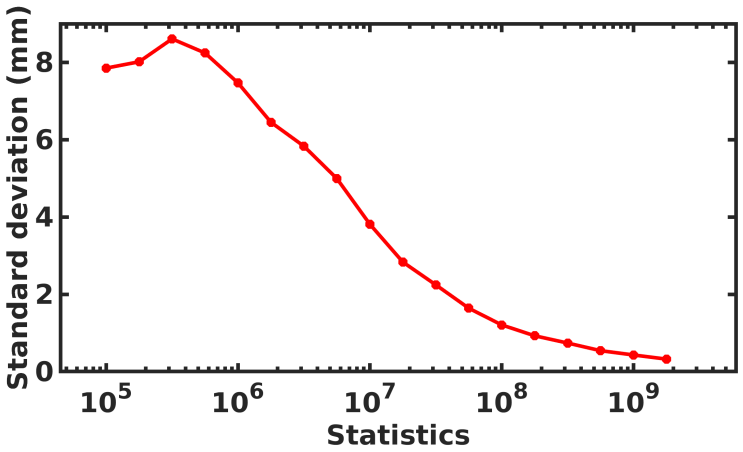


Figure 6.6: Standard deviation of the shifts between PG measurement profiles and the corresponding PG simulation profiles under different PB statistics. Note that PG measurement profiles below 5.9×10^5 may be invalid due to low PG counting as explained in figure 6.5 and the text.

and a std of 0.9 mm. The 7 failure cases, exhibiting maximum deviation of 3.8 mm, are distributed at the edge of the recommended area.
For the PG simulation, there are 21 PBs that are recommended by the PG emission indicator but have shift differences higher than 1.5 mm. Applying the camera indicator, 10 of those failure cases are rejected because they are not centered at the camera FOV. Only 2 spots of the remaining 11 PBs have shift difference values higher than 2 mm, with a maximum value of 4.5 mm. These two spots are among the above mentioned 7 failure cases of PG emission indicator (an example has been shown in the figure 6.8).
To compare the methods of PB boosting and PB aggregation (with a sigma of 7 mm, see section 5.2) at the detection level, three PBs are selected among the PBs which are (1) recommended by both indicators, (2) recommended only by the PG emission indicator, (3) recommended only by the camera indicator, (4) rejected by all indicators (see figure 6.9). The results are shown in table 6.1. For the PBs recommended by both indicators, the maximum shift difference values are 1.6 mm (aggregation) and 1.4 mm (boosting) with mean values of 0.8 mm (aggregation) and 0.5 mm (boosting) along with std of 0.7 mm (aggregation) and 0.6 mm (boosting). For the PBs recommended by only the PG emission indicator, the maximum shift differences are -1.9 mm (aggregation) mm and 3.8 mm (boosting) with mean values of 1.0 mm (both) along with std of 1.0 mm (aggregation) mm and 1.6 mm (boosting). For the PBs recommended by only the camera indicator, the maximum shift differences are -5.6 mm (aggregation) and 5.1 mm (boosting) with mean values of 1.0 mm (aggregation) and 1.1 mm (boosting) along with std of 2.8 mm (aggregation) mm and 2.6 mm (boosting). For the PBs rejected by all the indicators, the maximum shift differences are 13.2 mm (aggregation) and 9.1 (boosting) with mean values of 1.4 mm (aggregation) and 0.6 mm (boosting) along with std of 4.7 mm (aggregation) mm and 4.5

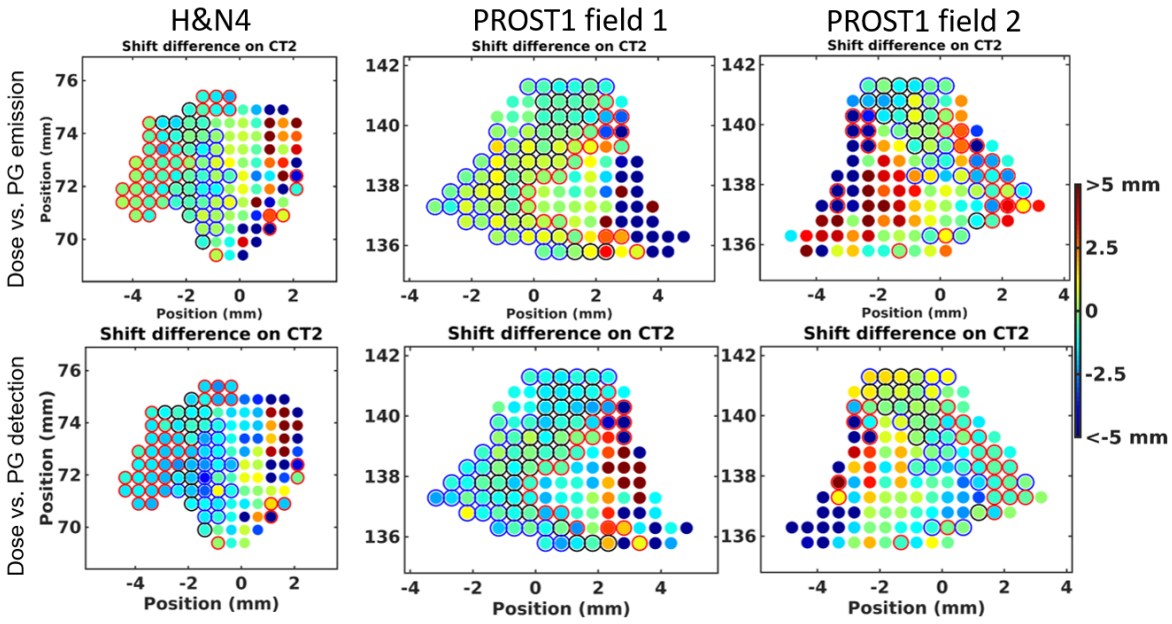


Figure 6.7: The shift differences between dose and PG emission are shown on the top and the shift differences between dose and PG simulation of the ideal detected signal are shown on the bottom for the three considered fields (left to right). The shift difference is color coded. PBs with black circles are recommended by both PG emission and camera indicators. PBs with red circles are recommended only by the camera indicator, while PBs with blue circles are recommended only by the PG emission indicator.

mm (boosting).

6.3.3 Camera response scenario of comparing the measurement on CT2 to the simulation on CT1

The shifts between the generated 1000 PG measurement profiles on CT2 and the corresponding PG simulation profiles on CT1 are compared to the shifts of the dose profiles on different CTs. The results, quantified as the mean and std of the shift differences, are shown in the middle and bottom rows in figure 6.10 using statistics from the initial CERR TP and boosted statistics, respectively. And the ground truth, i.e. shift difference using the PG simulation profiles, is given on the top as reference. The shift differences calculated using the high statistics scenario approach the ground truth with a mean deviation of 0.2 mm for all PBs. The deviations of all PBs recommended by both indicators are lower than 2 mm. Most of the PBs in the initial TP do not have enough protons for valid PG detection.

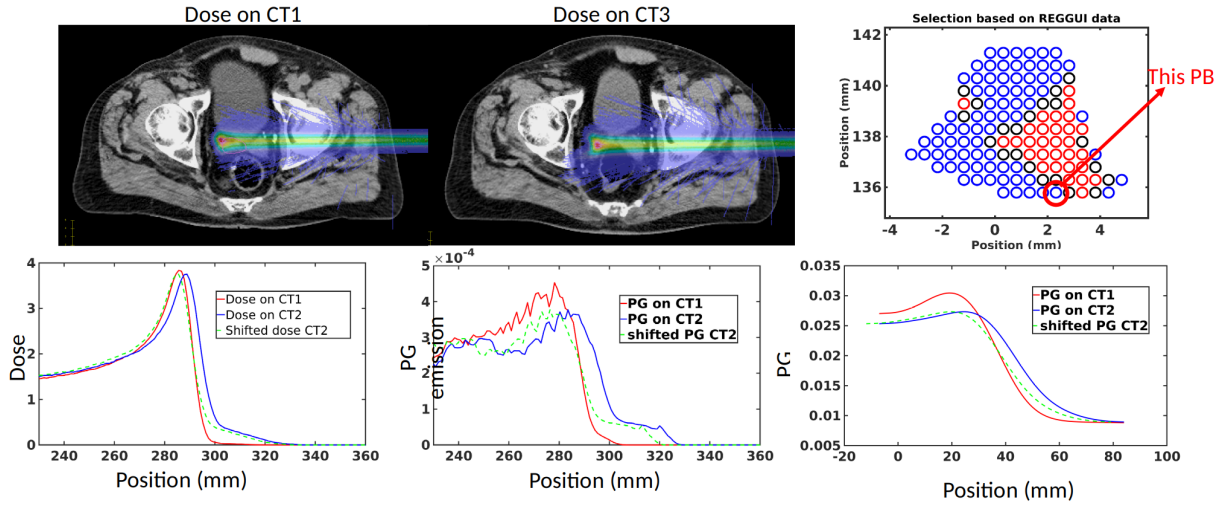


Figure 6.8: The magnitude of the change of the air cavity near the beam path of this PB is bigger than the safe margin set for the indicators (cf. section 5.5.3), leading to a failure of the PG emission indicator. The position of this PB in BEV and the axial slice of the 3D dose distribution on both CTs are shown on the top right and left, respectively. The shifts of the dose (bottom left), PG emission (bottom middle) and PG detection (bottom right) are 3.4, 7.2, and 4.8 mm, respectively.

Comparisons between PB boosting and PB aggregation approaches for the selected PBs are shown in table 6.2 (only statistical uncertainty is reported not accounting for uncertainty caused by e.g. the shifting step size, cf. section 5.5.1.). For the PBs recommended by both indicators, the maximum shift differences are 2.4 mm (aggregation) and 1.8 mm (boosting) with mean values of 1.3 mm (aggregation) and 0.8 mm (boosting) along with std of 1.0 mm (both). For the PBs recommended by only the PG emission indicator, the maximum shift differences are 2.7 mm (aggregation) mm and 3.9 mm (boosting) with mean values of 1.2 mm (aggregation) and 1.5 mm (boosting) along with std of 1.6 mm (both). For the PBs recommended by only the camera indicator, the maximum shift differences are 5.3 mm (aggregation) and 5.9 mm (boosting) with mean values of 2.0 mm (both) along with std of 2.2 mm (aggregation) and 2.3 mm (boosting). For the PBs rejected by the indicators, the maximum shift differences are 14.8 mm (aggregation) and 11.8 (boosting) with mean values of 3.5 mm (aggregation) and 3.2 mm (boosting) along with std of 5.0 mm (aggregation) and 4.3 mm (boosting).

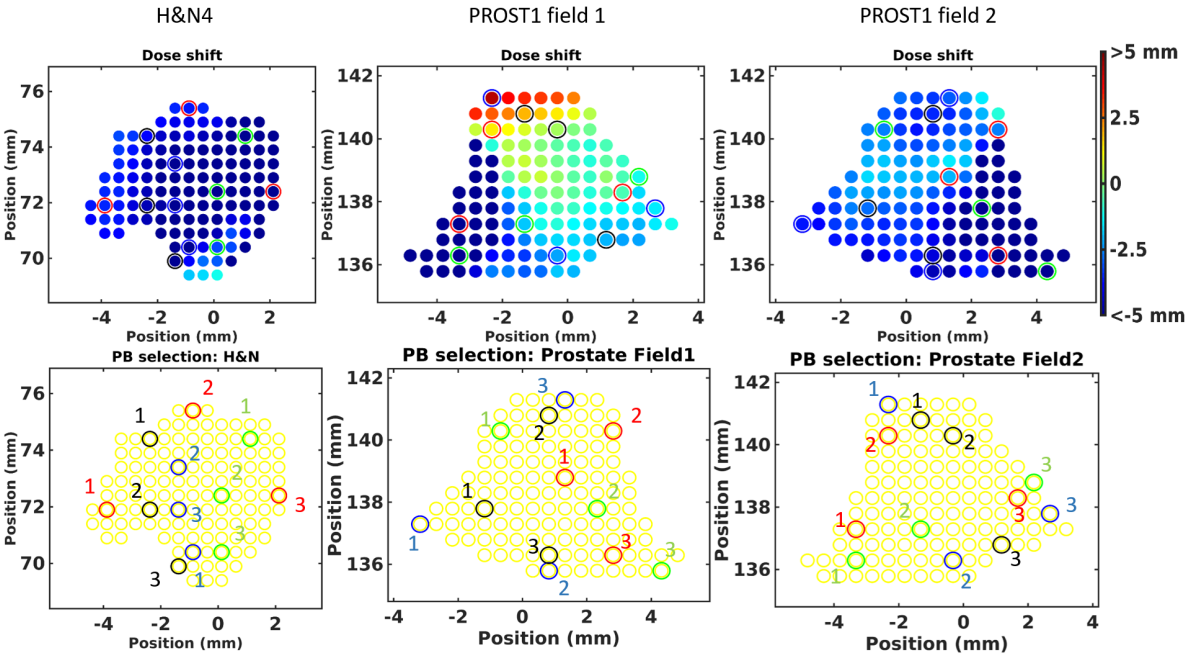


Figure 6.9: Figures on the top show the color coded dose shifts of the PBs investigated in this work, with circles referring to the PBs selected for the comparison between PB boosting and PB aggregation (see text). The positions are shown on the bottom. The spots with black, blue and red circles are those recommended by both indicators, only PG emission indicator and only camera indicators, respectively. PBs rejected by both indicators are with green circles.

6.3.4 Camera response scenario of comparing the PG measurements on different CTs

The shifts between the generated 1000 pairs of PG measurement on different CTs are compared to the shifts of the dose profiles on different CTs (figure 6.11). The results are shown in the middle and bottom rows using statistics from the initial CERR TP and boosted statistics, respectively. The shift differences using PG simulation profiles are shown on the top as ground truth. The shift differences calculated using high statistics approach the ground truth with a mean deviation of -1.3 mm for all PBs. The deviations of all PBs recommended by both indicators are lower than 2 mm. Most of the PBs in the initial TP do not have enough protons for valid PG detection.
Table 6.3 compares the PB boosting and PB aggregation approaches for the selected PBs in terms of the shift difference between dose and PG. Only statistical uncertainty is shown as mentioned above. For the PBs recommended by both indicators, the maximum shift differences are 2.4 mm (aggregation) and 2.0 mm (boosting) with mean values of 1.3 mm (aggregation) and 0.9 mm (boosting) along with std of 1.0 mm (aggregation) and 0.9 mm

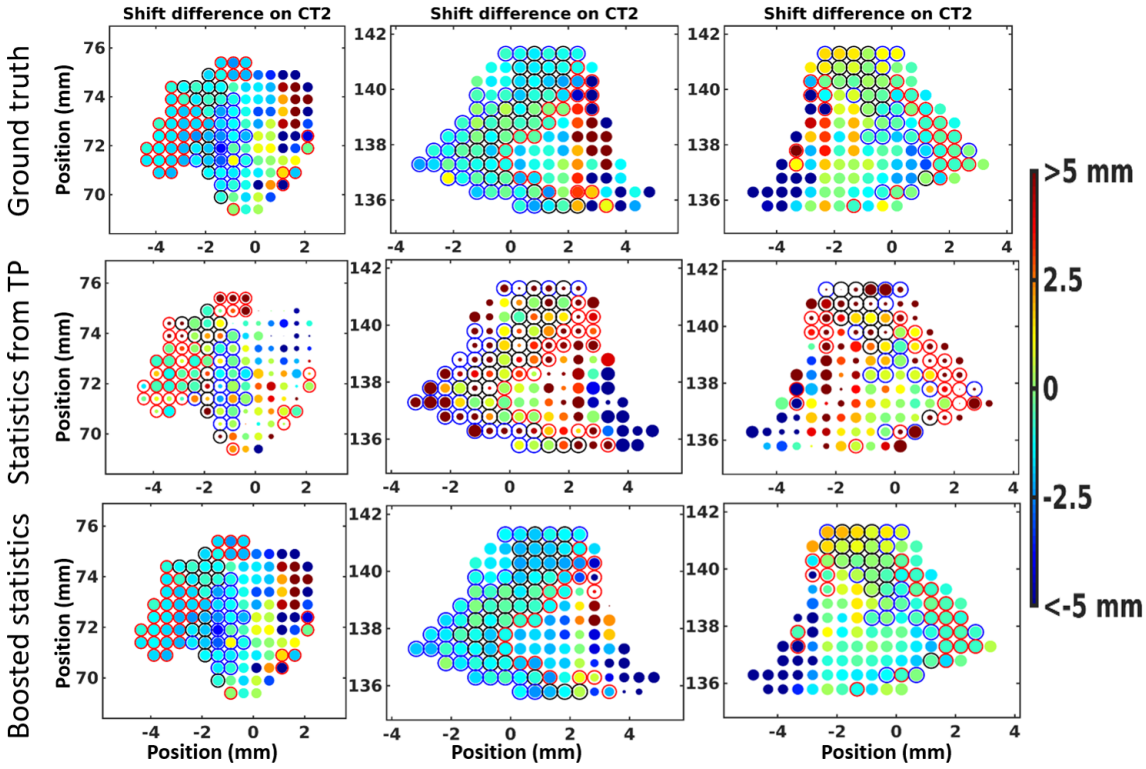


Figure 6.10: The shift differences between dose and PG simulation are shown on the top. The shift differences between dose and PG measurement (measurement vs. simulation) using the initial TP statistics and the boosted statistics are shown in the middle and bottom, respectively. The mean shift difference is color coded and the spot size indicates the std of the shift values, i.e. bigger spots have smaller std. PBs with black circles are those recommended by both PG emission and camera indicators. PBs with red circles are recommended by only the camera indicator and PBs with blue circles are recommended by only the PG emission indicator.

(boosting). For the PBs recommended by only the PG emission indicator, the maximum shift differences are 2.6 mm (aggregation) mm and 4.1 mm (boosting) with mean values of 1.4 mm (aggregation) and 1.6 mm (boosting) along with std of 1.5 mm (aggregation) and 1.8 mm (boosting). For the PBs recommended by only the camera indicator, the maximum shift differences are 6.1 mm (aggregation) and 5.7 mm (boosting) with mean values of 2.3 mm (both) along with std of 2.4 mm (aggregation) and 2.2 mm (boosting). For the PBs rejected by the indicators, the maximum shift differences are 15.8 mm (aggregation) and 11.8 (boosting) with mean values of 3.6 mm (aggregation) and 3.3 mm (boosting) along with std of 5.0 mm (aggregation) and 4.1 mm (boosting).

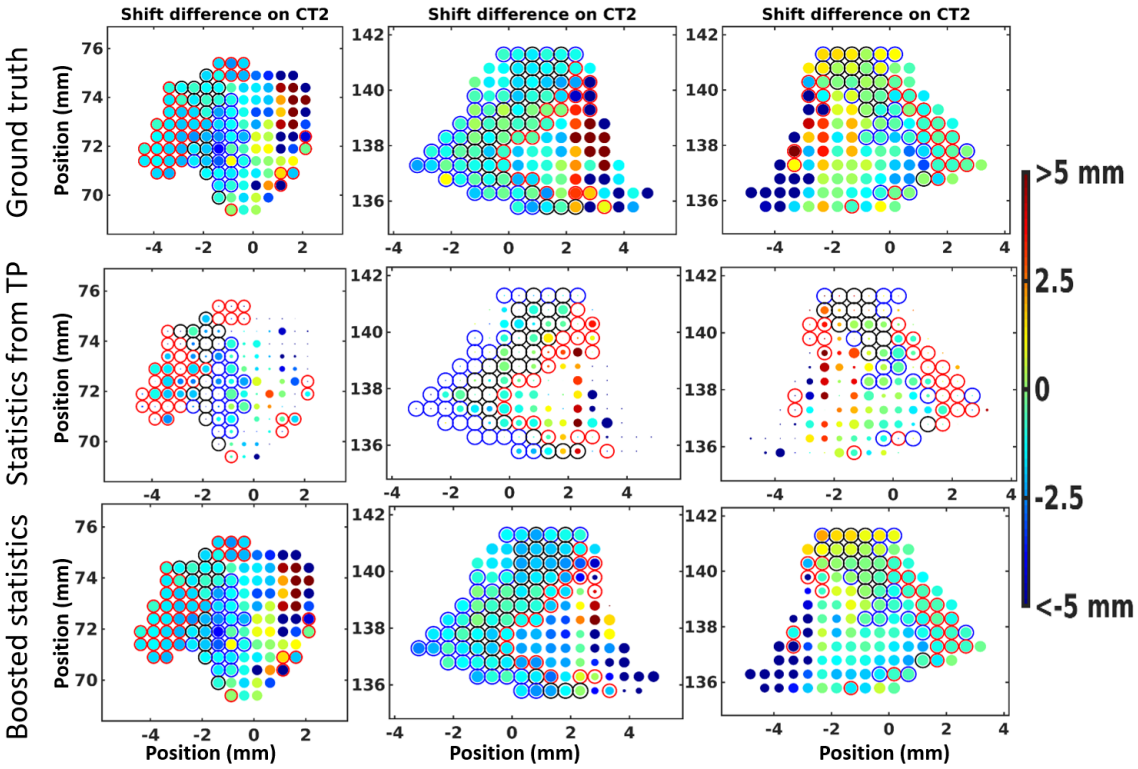


Figure 6.11: The shift differences between dose and PG simulation are shown on the top. The shift differences between dose and PG measurement (measurement vs. measurement) using the initial TP statistics and the boosted statistics are shown in the middle and bottom, respectively. The mean shift difference is color coded and the spot size indicates the std of the shift values, i.e. bigger spots have smaller std. PBs with black circles are those recommended by both PG emission and camera indicators. PBs with red circles are recommended by only the camera indicator and PBs with blue circles are recommended by only the PG emission indicator.

6.3.5 Statistics comparison between the PB aggregation and PB boosting

The statistics, range mixing and contribution of the central beam to the aggregated PBs (see section 5.7) are shown in the table 6.4. Statistics of 11 aggregated PBs are observed to be lower than 1×10^8. Two of them (prostate field2 PB 1 and 2) are recommended by both indicators and have higher PG shift error (1.6 mm for the comparison between measurement and simulation and more than 2.3 mm for the comparison between measurements on different CTs) than other recommended PBs (cf. table 6.2 and 6.3). The maximum range mixing observed is 12.1 mm, and the general contribution of the central PBs is low (with a maximum value of 57.1%). These results are consistent to the previous findings in section 5.7 though different PBs are selected.

6.4 Discussion

In this chapter, an indicator proposed by (Huang 2020) accounting for the effect of the relative position of the detector to the PG falloff is investigated for a knife-edge PG camera of limited FOV, which has been already deployed clinically by (Xie et al. 2017, Richter et al. 2016). PBs whose PG emission profile are supposed to reflect correct dose information but cannot be detected correctly by the camera are identified and rejected by the new indicator, which thus augments the work of chapter 5 to account for a realistic PG detector. Moreover, the influence of the camera response on the proposed boosting approach in comparison to the initial TP for a reliable range monitoring representative of the dose delivery is investigated for different PG detection scenarios under the different underlying statistics. Comparisons between the PB boosting and the alternative PB aggregation methods are also presented.

The PBs of interest for selection in this study, according to the method highlighted in chapter 5, are the maximum energy PBs which have their Bragg peaks inside the safe region in the PTV. The ranges of those PBs are varying, depending on the shape of the PTV. Thus, only some of the PBs of interest have their PG falloff centered in the FOV of the camera for a given camera position. In this work, the camera is set to the position where the number of the centered PBs is maximized. Another option could be to set candidate PBs for boosting to those which have their PG falloffs at the same depth, depending on the users interest.

For the investigated clinical cases of a head and neck and prostate cancer patient, the initial TPs do not provide sufficient statistics for most of the PBs to generate enough PG detection signal. The simulation based on the REGGUI platform shows that a precision of 0.8 mm can be achieved when PBs have more than 2×10^8 protons for the considered PG camera. This is consistent to previous studies with the same computational platform (Xie et al. 2017) or Geant4 for an ideal detector (Tian et al. 2018). All investigated scenarios of emulated measurement at a certain day versus the simulation on the planning CT and the comparison between emulated measurements at different fractions show that the information provided by the emulated PG measurement approaches the ground truth, i.e. the simulated PG detected signal without noise, with increasing statistical weight (i.e., number of protons) of the PBs. Besides, the general precision of the scenario of emulated measurement vs. simulation is observed to be better than that of measurement vs. measurement, as the result of the Poisson noise introduced in the measurement at the first time point. This thus supports the currently mostly implemented PG monitoring approach, comparing a given daily measurement to the prediction obtained on the basis of the treatment plan, provided that a reliable calculation engine is used for such a prediction.

The maximum shift difference between dose and PG on different CTs observed in ideal cases (simulation vs. simulation) for the PBs recommended by both the emission indicators of chapter 5 and the new camera position indicator are 1.9 mm and 1.5 mm for

PB aggregation and PB boosting with mean values of 0.8 mm and 0.5 mm, respectively. Both approaches are comparable. However, PB aggregation does not guarantee enough statistics for the aggregated PBs. In this work, the minimum aggregation statistics is 3.6×10^7 and there are 11 PBs that have less than 1.0×10^8 protons among the 24 selected PBs including 2 PBs recommended by both indicators. These two PBs have larger PG shift errors (1.6 mm for the comparison between measurement and simulation and more than 2.3 mm for the comparison between measurements on different CTs), compared to other recommended PBs, leading to a higher uncertainty in proton range verification. As demonstrated in previous chapters, the PB boosting approach shows advantages in terms of guaranteed statistics, range mixing, lateral resolution and the contribution of the central PB but there is no conflict between those two approaches.
PBs recommended by only the camera indicator are centered in the FOV of the camera. Although these PBs are not recommended by the PG emission indicator, their PG signal can still correctly reflect the PG emission. Thus, these PBs can still provide valid information to trigger additional investigations, e.g. a new CT as discussed in chapter 5. For PBs rejected by the camera indicator, their PG detection signals are not well correlated with the corresponding PG emission and cannot be used for proton range verification, regardless whether they are recommended by the PG emission indicator or not.
Failure cases of the PG emission indicator are observed for a few recommended PBs that exhibit large (up to 3.8 mm) shift difference between the dose and PG emission. These PBs are distributed at the edge of the recommended PB area. A safety ring of 5 mm (cf. section 5.5.4) is applied for the PG emission indicator. For larger anatomical changes, the indicator might fail. Besides, the PG emission indicator is developed based on the dataset generated by Geant4. The profiles generated by REGGUI are slightly different from those generated by Geant4. This causes different PB identification as shown in figure 6.3. A refinement of the algorithm or machine learning model trained using PG detection/emission generated by REGGUI simulation of experimental data could be helpful to further improve this process. The camera indicator could also fail due to large anatomical changes, e.g. the Bragg peak is at the center of FOV on CT1 but moves outside of the camera FOV on CT2 due to anatomical changes. In this case, a large PG shift can be observed. Although PG detection data is available in this work, the PG dose correlation is still quantified using PG emission. On the one hand, PG emission provides more explicit information of the effect of tissue heterogeneity, which is an important source of the mismatching between PG and dose, not losing information due to the PG transport in tissue until eventual detection. On the other hand, the PG emission information is independent of the camera type and needs less computational resources, compared to those needed for the other camera-free phsp data (see section 5.1), thus providing the feasibility to be combined with other camera specific indicators.
Besides few exceptions which could be overcome in the future by e.g. fine tuning of the algorithm for camera specific PG signal, avoidance of PBs in the edge areas and better safety margin setting, it could be generally shown that the proposed boosting approach with emission selection criteria complemented by the camera position can still work, i.e., provide a more reliable PG monitoring than for the initial TP plan, when evaluated at

detection considering a realistic PG camera respons

Patient & PB ID	ΔDose	ΔPG (aggregation)	ΔShift (aggregation)	ΔPG (boost)	ΔShift (boost)
PB recommended by both indicators					
H&N4 field1 1	3.7	2.1	1.6	3.1	0.6
H&N4 field1 2	4.4	2.9	1.5	3.0	1.4
H&N4 field1 3	4.9	3.9	1.0	4.8	0.1
PROST1 field1 1	2.6	2.0	0.6	2.1	0.5
PROST1 field1 2	3.3	1.9	1.4	2.1	1.3
PROST1 field1 3	4.0	3.0	1.0	3.3	0.8
PROST1 field2 1	-2.1	-1.9	-0.2	-1.8	-0.3
PROST1 field2 2	-0.4	-1.0	0.6	-0.8	0.4
PROST1 field2 3	1.8	2.1	-0.2	2.1	-0.3
PB recommended by only PG emission indicators					
H&N4 field1 1	4.7	3.0	1.7	4.0	0.7
H&N4 field1 2	5.1	3.9	1.2	3.0	2.1
H&N4 field1 3	4.9	3.1	1.9	1.1	3.8
PROST1 field1 1	3.6	1.8	1.8	1.5	2.0
PROST1 field1 2	4.4	3.2	1.1	3.0	1.4
PROST1 field1 3	2.7	1.2	1.5	2.6	0.1
PROST1 field2 1	-4.6	-3.1	-1.4	-2.9	-1.7
PROST1 field2 2	2.0	0.4	1.6	1.1	0.9
PROST1 field2 3	1.4	1.1	0.3	1.5	-0.1
PB recommended by only camera indicator					
H&N4 field1 1	3.3	2.1	1.1	2.1	1.2
H&N4 field1 2	3.4	2.0	1.4	1.1	2.3
H&N4 field1 3	7.0	2.8	4.3	3.0	4.1
PROST1 field1 1	1.9	0.8	1.1	0.4	1.5
PROST1 field1 2	2.4	-3.2	5.6	-2.7	5.1
PROST1 field1 3	5.9	9.4	-3.6	9.3	-3.5
PROST1 field2 1	26.1	27.7	-1.6	27.4	-1.4
PROST1 field2 2	-1.5	-1.4	-0.0	-1.4	-0.0
PROST1 field2 3	0.4	0.1	0.3	0.2	0.3
PB rejected by both indicators					
H&N4 field1 1	8.2	11.4	-3.2	14.2	-6.0
H&N4 field1 2	5.6	5.2	0.4	7.5	-1.9
H&N4 field1 3	2.9	3.8	-0.8	-3.0	6.0
PROST1 field1 1	2.4	1.1	1.3	1.7	0.7
PROST1 field1 2	11.3	11.6	-0.4	13.4	-2.1
PROST1 field1 3	23.6	21.7	1.9	23.3	0.3
PROST1 field2 1	18.9	5.7	13.2	9.8	9.1
PROST1 field2 2	1.8	1.9	-0.1	2.8	-1.0
PROST1 field2 3	1.0	0.7	0.3	0.6	0.4

Table 6.1: Comparison between the shifts of the dose (ΔDose) and ideal detected PG profiles (ΔPG for simulation vs. simulation) on different CTs for PB aggregation and boosting approaches (see location of selected spots in figure 6.9). The unit for the following values are mm. The worst shift difference in each cases are highlighted in red.

Patient & PB ID	ΔDose	ΔPG (aggregation)	Mean Δshift (aggregation)	ΔPG (boost)	Mean Δshift (boost)
PB recommended by both indicators					
H&N4 field1 1	3.7	1.7 ± 0.7	2.1	2.6 ± 0.9	1.1
H&N4 field1 2	4.4	2.2 ± 0.9	2.2	2.6 ± 0.9	1.8
H&N4 field1 3	4.9	2.5 ± 1.0	2.4	3.1 ± 1.0	1.8
PROST1 field1 1	2.6	1.6 ± 1.1	1.0	1.6 ± 1.5	1.0
PROST1 field1 2	3.3	1.3 ± 1.2	2.0	1.7 ± 1.5	1.7
PROST1 field1 3	4.0	2.3 ± 1.5	1.7	3.2 ± 1.6	0.8
PROST1 field2 1	-2.1	-1.5 ± 1.6	-0.6	-1.0 ± 1.3	-1.2
PROST1 field2 2	-0.4	-0.9 ± 1.6	0.5	-0.3 ± 1.3	-0.1
PROST1 field2 3	1.8	1.5 ± 1.4	0.3	1.7 ± 1.5	0.1
PB recommended by only PG emission indicators					
H&N4 field1 1	4.7	2.1 ± 0.7	2.6	2.9 ± 1.1	1.9
H&N4 field1 2	5.1	2.4 ± 0.9	2.7	2.2 ± 1.0	2.8
H&N4 field1 3	4.9	2.6 ± 0.8	2.3	1.1 ± 0.9	3.9
PROST1 field1 1	3.6	2.2 ± 1.9	1.3	1.7 ± 1.5	1.8
PROST1 field1 2	4.4	2.6 ± 1.4	1.8	2.1 ± 1.5	2.2
PROST1 field1 3	2.7	1.5 ± 1.3	1.2	1.6 ± 1.4	1.1
PROST1 field2 1	-4.6	-2.2 ± 1.6	-2.4	-2.6 ± 1.4	-2.0
PROST1 field2 2	2.0	1.1 ± 1.9	0.9	0.9 ± 1.6	1.1
PROST1 field2 3	1.4	1.1 ± 1.8	0.3	0.8 ± 1.6	0.6
PB recommended by only camera indicator					
H&N4 field1 1	3.3	1.8 ± 0.6	1.4	1.7 ± 0.9	1.6
H&N4 field1 2	3.4	1.5 ± 0.7	1.9	1.1 ± 0.8	2.3
H&N4 field1 3	7.0	1.7 ± 1.2	5.3	2.3 ± 1.1	4.7
PROST1 field1 1	1.9	0.5 ± 2.1	1.4	0.9 ± 2.2	1.0
PROST1 field1 2	2.4	-2.0 ± 3.9	4.4	-1.9 ± 3.3	4.3
PROST1 field1 3	5.9	5.8 ± 4.6	0.0	6.0 ± 6.3	-0.1
PROST1 field2 1	26.1	20.7 ± 2.9	5.3	20.2 ± 2.0	5.9
PROST1 field2 2	-1.5	-1.4 ± 1.7	-0.0	-1.2 ± 2.0	-0.3
PROST1 field2 3	0.4	0.4 ± 2.3	0.0	-0.1 ± 1.6	0.5
PB rejected by both indicators					
H&N4 field1 1	8.2	5.3 ± 1.0	2.9	7.6 ± 0.9	0.6
H&N4 field1 2	5.6	3.3 ± 1.3	2.3	4.7 ± 1.2	0.9
H&N4 field1 3	2.9	2.6 ± 0.9	0.4	-2.3 ± 0.9	5.2
PROST1 field1 1	2.4	0.7 ± 2.1	1.7	0.9 ± 1.5	1.5
PROST1 field1 2	11.3	10.8 ± 6.1	0.4	11.4 ± 5.2	-0.1
PROST1 field1 3	23.6	15.0 ± 4.9	8.6	15.4 ± 8.0	8.2
PROST1 field2 1	18.9	4.1 ± 3.5	14.8	7.1 ± 3.9	11.8
PROST1 field2 2	1.8	1.6 ± 1.5	0.2	1.7 ± 2.1	0.1
PROST1 field2 3	1.0	0.9 ± 1.8	0.1	0.5 ± 1.6	0.5

Table 6.2: Comparison between the shifts of the dose (ΔDose) and PG detection profiles (ΔPG for measurement vs. simulation) on different CTs for PB aggregation and boosting approaches (see location of selected spots in figure 6.9). The unit for the following values are mm. The worst shift difference in each cases are highlighted in red.

Patient & PB ID	ΔDose	ΔPG (aggregation)	Mean Δshift (aggregation)	ΔPG (boost)	Mean Δshift (boost)
PB recommended by both indicators					
H&N4 field1 1	3.7	1.9 ± 0.9	1.8	2.7 ± 1.3	1.1
H&N4 field1 2	4.4	2.1 ± 1.2	2.2	2.4 ± 1.3	2.0
H&N4 field1 3	4.9	2.5 ± 1.3	2.4	3.3 ± 1.5	1.6
PROST1 field1 1	2.6	1.6 ± 1.6	1.1	1.4 ± 2.1	1.2
PROST1 field1 2	3.3	1.5 ± 1.7	1.9	1.6 ± 2.2	1.7
PROST1 field1 3	4.0	2.1 ± 2.1	1.9	2.9 ± 2.4	1.2
PROST1 field2 1	-2.1	-1.4 ± 2.4	-0.7	-1.2 ± 2.0	-1.0
PROST1 field2 2	-0.4	-0.8 ± 2.3	0.4	-0.4 ± 1.9	0.1
PROST1 field2 3	1.8	1.4 ± 2.0	0.5	1.5 ± 2.2	0.4
PB recommended by only PG emission indicators					
H&N4 field1 1	4.7	2.3 ± 1.0	2.4	2.8 ± 1.6	2.0
H&N4 field1 2	5.1	2.5 ± 1.3	2.6	2.0 ± 1.5	3.0
H&N4 field1 3	4.9	2.3 ± 1.2	2.6	0.8 ± 1.4	4.1
PROST1 field1 1	3.6	1.2 ± 2.8	2.4	1.1 ± 2.0	2.5
PROST1 field1 2	4.4	2.7 ± 2.1	1.7	2.1 ± 2.1	2.3
PROST1 field1 3	2.7	1.4 ± 1.9	1.3	1.6 ± 2.0	1.1
PROST1 field2 1	-4.6	-2.5 ± 2.2	-2.1	-2.3 ± 2.0	-2.2
PROST1 field2 2	2.0	0.6 ± 2.6	1.4	0.9 ± 2.3	1.1
PROST1 field2 3	1.4	0.7 ± 2.8	0.7	0.7 ± 2.3	0.7
PB recommended by only camera indicator					
H&N4 field1 1	3.3	1.8 ± 0.9	1.4	1.7 ± 1.3	1.6
H&N4 field1 2	3.4	1.6 ± 1.0	1.8	1.2 ± 1.2	2.2
H&N4 field1 3	7.0	1.8 ± 1.8	5.3	2.2 ± 1.6	4.8
PROST1 field1 1	1.9	0.1 ± 3.1	1.8	0.2 ± 3.5	1.7
PROST1 field1 2	2.4	-2.3 ± 5.4	4.7	-2.0 ± 4.6	4.4
PROST1 field1 3	5.9	6.2 ± 5.4	-0.4	5.7 ± 7.1	0.2
PROST1 field2 1	26.1	20.0 ± 4.7	6.1	20.3 ± 3.0	5.7
PROST1 field2 2	-1.5	-1.7 ± 2.5	0.2	-1.3 ± 2.9	-0.2
PROST1 field2 3	0.4	0.3 ± 3.3	0.2	0.1 ± 2.2	0.4
PB rejected by both indicators					
H&N4 field1 1	8.2	5.2 ± 1.5	3.0	7.5 ± 1.3	0.7
H&N4 field1 2	5.6	3.3 ± 1.8	2.3	4.6 ± 1.9	1.0
H&N4 field1 3	2.9	2.5 ± 1.3	0.4	-2.2 ± 1.5	5.2
PROST1 field1 1	2.4	0.9 ± 3.1	1.5	1.0 ± 2.2	1.4
PROST1 field1 2	11.3	10.9 ± 6.9	0.4	10.8 ± 6.2	0.5
PROST1 field1 3	23.6	15.4 ± 5.4	8.3	15.9 ± 8.1	7.7
PROST1 field2 1	18.9	3.7 ± 4.7	15.2	7.1 ± 4.7	11.8
PROST1 field2 2	1.8	1.3 ± 2.4	0.5	1.5 ± 3.1	0.3
PROST1 field2 3	1.0	0.2 ± 2.7	0.8	-0.1 ± 2.4	1.1

Table 6.3: Comparison between the shifts of the dose (ΔDose) and PG detection profiles (ΔPG for measurement vs. measurement) on different CTs for PB aggregation and boosting approaches (see location of selected spots in figure 6.9). The unit for the following values are mm. The worst shift difference in each cases are highlighted in red.

Patient & PB ID	ΔRange (mm)	$N_{\text{aggregation}}$	N_{central}	C
PB recommended by both indicators				
H&N4 field1 1	6.5	2.2×10^8	2.0×10^7	8.9%
H&N4 field1 2	4.6	1.3×10^8	1.3×10^7	10.2%
H&N4 field1 3	4.6	1.5×10^8	4.1×10^5	0.3%
PROST1 field1 1	0.8	2.7×10^8	2.4×10^6	0.9%
PROST1 field1 2	3.3	2.0×10^8	4.3×10^7	21.1%
PROST1 field1 3	2.8	1.3×10^8	5.3×10^6	3.9%
PROST1 field2 1	0.6	7.6×10^7	2.2×10^7	2.9%
PROST1 field2 2	1.4	8.5×10^7	2.0×10^6	23.6%
PROST1 field2 3	3.4	1.3×10^8	1.2×10^7	0.9%
PBs recommended by only PG emission indicators				
H&N4 field1 1	2.9	3.0×10^8	3.4×10^6	1.1%
H&N4 field1 2	2.5	1.9×10^8	1.1×10^7	5.8%
H&N4 field1 3	2.5	1.9×10^8	8.4×10^5	0.4%
PROST1 field1 1	1.0	4.2×10^7	2.5×10^6	6.0%
PROST1 field1 2	1.8	1.7×10^8	2.5×10^7	14.7%
PROST1 field1 3	3.1	1.6×10^8	2.0×10^6	1.3%
PROST1 field2 1	0.9	9.3×10^7	1.7×10^6	1.9%
PROST1 field2 2	3.2	9.0×10^7	2.3×10^7	25.3%
PROST1 field2 3	5.6	6.9×10^7	8.5×10^5	1.2%
PBs recommended by only camera indicators				
H&N4 field1 1	8.4	2.9×10^8	2.9×10^7	10.2%
H&N4 field1 2	2.7	1.8×10^8	3.2×10^5	0.2%
H&N4 field1 3	7.2	1.4×10^8	6.8×10^5	0.5%
PROST1 field1 1	1.9	1.8×10^8	1.1×10^7	5.9%
PROST1 field1 2	2.2	1.5×10^8	8.6×10^7	57.1%
PROST1 field1 3	8.2	1.2×10^8	4.2×10^6	3.5%
PROST1 field2 1	3.5	3.8×10^7	2.1×10^6	5.6%
PROST1 field2 2	0.5	1.4×10^8	2.8×10^7	20.4%
PROST1 field2 3	3.4	4.2×10^7	3.3×10^7	7.8%
PBs recommended by only camera indicators				
H&N4 field1 1	12.1	1.9×10^8	4.6×10^7	24.2%
H&N4 field1 2	7.8	2.0×10^8	3.6×10^7	18.3%
H&N4 field1 3	11.7	2.4×10^8	1.2×10^6	0.5%
PROST1 field1 1	1.8	3.6×10^7	1.3×10^6	3.5%
PROST1 field1 2	5.2	1.5×10^8	3.9×10^7	25.4%
PROST1 field1 3	4.0	1.8×10^8	1.3×10^6	0.7%
PROST1 field2 1	8.5	5.1×10^7	4.8×10^6	9.5%
PROST1 field2 2	4.2	2.7×10^8	4.1×10^7	15.4%
PROST1 field2 3	4.3	9.5×10^7	4.6×10^7	48.8%

Table 6.4: Range mixing, accumulated statistics of the aggregated PBs, statistics of the central PB in the initial TP and its corresponding percentage contribution to the aggregated signal are given by ΔRange, $N_{aggregation}$, $N_{central}$ and c, respectively

Chapter 7

Conclusion and outlook

The clinical potential of proton therapy highly relies on the dose delivery precision. However, due to e.g. the semi-empirical conversion of the X-ray CT numbers into tissue stopping power ratio (relative to water) along with patient positioning and anatomical changes, the proton range in the patient has uncertainties. Prompt gamma is currently one of the most promising technologies providing the possibility for in vivo and ideally real-time proton range verification. The accuracy of PG is affected by tissue heterogeneities, potential inter- (and, although not addressed in this work, intra-) fractional anatomical changes as well as the position of the PG detector relative to the patient, especially in the case of limited field-of-view systems like the first clinical prototype of a single slit knife-edge camera. Moreover, detectability and reproducibility of the PG signal critically depends on sufficient counting statistics, which is generally not guaranteed in a conventional treatment plan.
In this work, a so called PG emission indicator, consisting of the PG-dose correlation and dose surface indicators, is proposed at the PG emission and treatment planning level to identify the conformities between the PG and dose profiles as well as the sensitivities of the PG-dose correlation to tissue heterogeneities. Additionally, the camera position indicator proposed by (Huang 2020) is investigated to account for a realistic detector response, including effects due to the counting statistics and the relative position of the camera and PG signal falloff. All these indicators are tested on PBs of interest in the treatment plan for four head and neck and one prostate cancer patients using both Geant4 and REGGUI generated dose and PG profiles. For the PG emission indicator, the shift differences between the dose and PG emission distributions of the recommended PBs evaluated on CTs at different time points (to capture anatomical changes) are within 1 mm for Geant4 datasets and less than 1.5 mm for 90% PBs simulated by REGGUI (with exceptions mostly due to the application of the algorithm and machine learning model constructed based on Geant4 datasets). The camera position indicator rejects PBs that are not centered in the FOV of the above mentioned camera. Combining both indicators, only 2 PBs , located at the edge of the area of the recommended PBs, were found to have more than 2 mm (maximum 4.5 mm) of difference between the falloff shifts of dose and detected/emitted PG on the different CTs (REGGUI), due to comparably large anatomical changes at the Bragg peak. Besides a few exceptions caused by unpredictable and comparably large anatomi-

cal changes, the PG signals of most of the PBs recommended by the proposed indicators are proven to be a reliable surrogate of the dose falloff at emission/detection level on the same/different CTs, thus being suitable for proton range monitoring. Hence, the introduced indicators are valuable to be used to identify and recommend reliable PBs for PG based proton range verification, provided that sufficient statistics is available for a reliable PG detection.

The Geant4 simulation used in this work shows that a precision of 1 mm in the PG signal detected by an ideal scoring plane outside the target (phase space data, see section 5.1) can be achieved at a PB proton statistics of 1.35×10^8, while the REGGUI simulation including the slit-camera response shows that a 0.8 mm precision can be achieved at a PB proton statistics of 2.00×10^8. These results are consistent to previous studies of (Xie et al. 2017, Draeger et al. 2018). Only a few PBs can reach such a high statistics in conventional TPs for typical fractionated proton treatments. To overcome this limitation, in this work a method for boosting a few selected PBs above the statistics threshold while not spoiling the TP was proposed and compared to other alternative approaches suggested in the literature such as PB aggregation (Xie et al. 2017), which aggregates nearby PBs of the initial plan to achieve high statistics. The re-optimized TPs are proven equivalent to the initial TPs in terms of dose distribution, dose averaged LET distribution and TP robustness with respect to anatomical changes, translational changes as well as range variations. The PB boosting method shows advantages over aggregation in terms of guaranteed statistics, improved lateral resolution and reduced range mixing. Moreover, there is no conflict between both approaches, as PG data of individual PBs could still be aggregated regardless whether PBs are boosted or not.

With the results of this thesis, the in-vivo PG range verification method is taken into account in the MC proton TP for the first time, potentially contributing to a future reduction of range uncertainties.

PG based proton range verification is based on the detected PG signal, which highly depends on the PG camera. In this work, the camera response of a knife-edge slit camera, which has been the first PG detector finding its way to clinical application, has been discussed. Further research could include the validation against other camera models, e.g. Compton camera. This latter technology provides the possibility to monitor PG in 3D, thus the 3D proton range information is possible to be monitored in the future. The proposed approach in this thesis, so far applied to laterally integrated depth profiles, could be further extended to fully account for a detected 3D PG signal.

In this work, PBs are identified and classified into three categories: 1. good PBs recommended by both PG emission indicator and camera position indicator, which in turn reflect precise proton range information; 2. counter-indicated PBs, which are centered at the FOV but do not allow to infer precisely the dose shift from the PG shift; 3. PBs which are not centered in the FOV of the camera. When selecting the PBs to boost in a TP, good PBs should be selected to provide reliable range information, e.g., for ruling out systematic errors in the CT-range calibration or set-up errors. The counter-indicated PBs are usually distributed in regions that are more sensitive to the anatomical changes, and could thus be selected to provide valuable information on unexpected range errors due to such changes,

although with less precise correlation to the delivered dose. More studies are needed to further investigate the amount of information provided by the PG signal in relation to the dose for the different categories, such that potential strategies, e.g. whether to trigger a new CT or not, can be made correctly.
The next step should consider integration of the proposed method in treatment planning platforms connected to treatment facilities equipped with PG detectors for a first experimental proof-of-concept that the method proposed in this thesis can indeed enable more reliable monitoring at both emission and detection level, e.g. in some controlled phantom irradiation, prior to a later clinical deploymen

Acknowledgements

First of all, I want to thank my supervisor, Prof. Katia Parodi. Thank you so much for giving me this chance to do the research for my PhD thesis in your group at LMU. It was with your encouragement and patient guidance that I can not only successfully conduct the research but also explore and improve my skills. I appreciate your ideas which play essential roles in this work, e.g. PB boosting, as well as your corrections for the writing of the paper and thesis etc. I'm fortunate to have you as my supervisor. I will always remember what you taught me which are much more than the research in this project. Thank you!

Apart from my supervisor I would say thanks to Prof. Carsten Welsch for organizing the OMA network, which provides not only the funding but also a series of valuable training through all the institutes involved. Your supports are essential to this work.

I want to also thank Prof. Belka as this work is based on the valuable clinical data provided by you.

Moreover, I would express my gratitude to Prof. Dr Guillaume Landry and Dr. Georgios Dedes for your teaching and support throughout my PhD. I am also pleased to thank Dr. Marco Pinto for your valuable and critical comments. It wouldn't have been possible to conduct this research smoothly and strictly without your precious helps.

Besides, I want to say thanks to Dr. Florian Kamp, Dr. Guillaume Janssens and my office mates Katharina Niepel, Jannis Dickmann and Ze Huang for helping me to get familiar with the tools used in this research as well as answering my questions patiently. Research are much easier with your help.

In the end, I acknowledge the funding supports from the he European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 675265, as well as the financial support from the DFG Excellence Cluster 'Munich Center for Advanced Photonics'.

Bibliography

Abolfath, R., Peeler, C., Newpower, M., Brink, L., Grosshans, D. & Mohan, R. (2017). A model for relative biological effectiveness of therapeutic proton beams based on a global fit of cell survival data, *Scientific Reports* p. 8340.

Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., Banerjee, S., Barrand, G., Behner, F., Bellagamba, L., Boudreau, J., Broglia, L., Brunengo, A., Burkhardt, H., Chauvie, S., Chuma, J., Chytracek, R. & Zschiesche, D. (2003). Geant4 - a simulation toolkit.

Ajzenberg-Selove, F. (1990). Energy levels of light nuclei a = 11-12, *Nuclear Physics A - NUCL PHYS A* **506(1)**: 1–158.

Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Barrand, G., Capra, R., Chauvie, S., Chytracek, R., Cirrone, P., Cooperman, G., Cosmo, G., Cuttone, G., Daquino, G., Donszelmann, M., Dressel, M., Folger, G., Foppiano, F. & Yoshida, H. (2006). Geant4 developments and applications, *IEEE Transactions on Nuclear Science* **53**: 270–278.

Allison, J., Amako, K., Apostolakis, J., Arce, P., Asai, M., Aso, T., Bagli, E., Bagulya, A., Banerjee, S., Barrand, G., Beck, B., Bogdanov, A., Brandt, D., Brown, J., Burkhardt, H., Canal, P., Ott, D., Chauvie, S., Cho, K. & Yoshida, H. (2016). Recent developments in geant4, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **835**.

Alpaydin, E. (2010). *Design and Analysis of Machine Learning Experiments*, MIT Press.

Andreo, P. (1991). Monte Carlo techniques in medical radiation physics, *Phys Med Biol* **36**: 861–920.

Andreo, P. & Benmakhlouf, H. (2017). Role of the density, density effect and mean excitation energy in solid-state detectors for small photon fields, *Phys Med Biol* **62**.

Assmann, W., Kellnberger, S., Reinhardt, S., Lehrack, S., Edlich, A., Thirolf, P., Moser, M., Dollinger, G., Omar, M., Ntziachristos, V. & Parodi, K. (2015). Ionoacoustic characterization of the proton bragg peak with submillimeter accuracy, *Journal of Medical Physics* **42**.

Baro, J., Sempau, J., Fernndez-Varea, J. & Salvat, F. (1995). Penelope: An algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter, *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **100**: 31–46.

Battistoni, G., Mattei, I. & Muraro, S. (2016). Nuclear physics and particle therapy, *Advances in Physics: X* **1**: 1–26.

Baumann, M., Krause, M., Overgaard, J., Debus, J., Bentzen, S. M., Daartz, J., Richter, C., Zips, D. & Bortfeld, T. (2016). Radiation oncology in the era of precision medicine, *Nat Rev Cancer* **16**(4): 234–49.

Bellinzona, V. E., Ciocca, M., Embriaco, A., Fontana, A., Mairani, A., Mori, M. & Parodi, K. (2015). On the parametrization of lateral dose profiles in proton radiation therapy, *Phys Med* **31**(5): 484–92.

Berger, M., Coursey, J., Zucker, M. & Chang, J. (2017). Stopping-power & range tables for electrons, protons, and helium ions, *Technical report, NIST* **NISTIR**: 4999.

Bertini, H., Gabriel, T., Santoro, R., Hermann, O., Larson, N. & Hunt, J. (1974). Hic1: a first approach to the calculation of heavy-ion reactions at energies greater than or equal to 50 mev/nucleon, *Technical report, Oak ridge national laboratory, Oak ridge Tennessee* .

Biegun, A. K., Seravalli, E., Lopes, P. C., Rinaldi, I., Pinto, M., Oxley, D. C., Dendooven, P., Verhaegen, F., Parodi, K., Crespo, P. & Schaart, D. R. (2012). Time-of-flight neutron rejection to improve prompt gamma imaging for proton range verification: a simulation study, *Phys Med Biol* **57**(20): 6429–6444.

Blann, H. (1983). Precompound analyses of spectra and yields following nuclear capture of stopped π^-, *Phys. Rev. C* **28**.

B"ohlen, T., Cerutti, F., Chin, M., Fasso, A., Ferrari, A., G. Ortega, P., Mairani, A., Sala, P., Smirnov, G. & Vlachoudis, V. (2014). The fluka code: Developments and challenges for high energy and medical applications, *Nuclear Data Sheets* **120**: 211–214.

Bom, V., Joulaeizadeh, L. & Beekman, F. (2011). Real-time prompt gamma monitoring in spot-scanning proton therapy using imaging through a knife-edge-shaped slit, *Phys Med Biol* **57**: 297–308.

Bortfeld, T. (2006). Imrt: a review and preview, *Phys Med Biol* **51**(13): R363–79.

Bortfeld, T. & Schlegel, W. (1996). An analytical approximation of depth-dose distributions for therapeutic proton beams, *Phys Med Biol* **41**: 1331–9.

Bragg, W. & Kleeman, R. (1904). Lxxiv. on the ionization curves of radium, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **8(48)**: 726–738.

Brown, F. & Martin, W. (2003). Direct sampling of Monte Carlo flight paths in media with continuously varying cross-sections.

Busca, P., Peloso, R., Fiorini, C., Gola, A., Abba, A., Erlandsson, K., Hutton, B., Bianchi, C., Poli, G., Guerra, U., Virotta, G., Ottobrini, L., Martelli, C., Lucignani, G., Pedretti, A., Mullekom, P., Incorvaia, S. & Perotti, F. (2010). Applications of the hicam gamma camera, pp. 2104 – 2107.

Chen, Y. & Ahmad, S. (2011). Empirical model estimation of relative biological effectiveness for proton beam therapy, *Radiation protection dosimetry* **149**: 116–23.

Das & Paganetti (2015). Introduction and history of proton therapy, *Principles and practice of proton beam therapy, AAPM Monograph2015, Summer School* **chapter 1**: pages 1–15.

Deasy, J. O., Blanco, A. I. & Clark, V. H. (2003). CERR: A computational environment for radiotherapy research, *Med Phys* **30**(5): 979–985.

Draeger, E., Mackin, D., Peterson, S., Chen, H., Avery, S., Beddar, S. & Polf, J. C. (2018). 3d prompt gamma imaging for proton beam range verification, *Phys Med Biol* **63**(3): 533–536.

Draeger, E., Mackin, D., Peterson, S., Chen, H., Beddar, S. & Polf, J. (2016). Su-f-j-189: A method to improve the spatial resolution of prompt gamma based compton imaging for proton range verification, *Med Phys* **43**: 3451–3451.

Everett, D., Fleming, J., Todd, W. & Nightingal, J. (1977). Gamma-radiation imaging system based on the compton effect, *Electrical Engineers, Proceedings of the Institution of Engineering and Technology* **124(11)**: 995.

Feeman, T. (2010). *The Mathematics of Medical Imaging: A Beginner's Guide*.

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Pineros, M., Znaor, A. & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods, *Int J Cancer* **144**(8): 1941–1953.

Fermi, E. (1950). High energy nuclear events, *Progress of Theoretical Physics* **5**: 570–583.

Ferrari, A. & Sala, P. (1998). The physics of high energy reactions, *Proceed. Workshop on Nucl. Reaction Data and Nuclear Reactor Physics, Design and Safety* **2**.

Ferrero, V., Fiorina, E., Morrocchi, M., Pennazio, F., Baroni, G., Battistoni, G., Belcari, N., Camarlinghi, N., Ciocca, M., Del Guerra, A., Donetti, M., Giordanengo, S., Giraudo, G., Patera, V., Peroni, C., Rivetti, A., Da Rocha Rolo, M., Rossi, S., Rosso, V. & Giuseppina, M. (2018). Online proton therapy monitoring: Clinical test of a silicon-photodetector-based in-beam pet, *Scientific Reports* **8**.

Fiorini, C., Gola, A., Peloso, R., Longoni, A., Lechner, P., Niculae, A., Soltau, H. & Strder, L. (2008). Silicon drift detectors arrays for the hicam gamma camera, pp. 2981 – 2983.

Frandes, M., Zoglauer, A., Maxim, V. & Prost, R. (2010). A tracking compton-scattering imaging system for hadron therapy monitoring, *IEEE Transactions on Nuclear Science* **57**(1): 144–150.

Frey, K., Unholtz, D., Bauer, J., Debus, J., Min, C. H., Bortfeld, T., Paganetti, H. & Parodi, K. (2014). Automation and uncertainty analysis of a method for in-vivo range verification in particle therapy, *Phys Med Biol* **59**(19): 5903–19.

Garfinkel-Castro, A., Bushman, T., Sabouri, S., Brewer, S., Song, Y. & Park, K. (2020). *Cluster Analysis*, pp. 121–153.

Geant4 (2013). Physics reference manual.

Geant4 (2018). Physics reference manual.

Golnik, C., Hueso-Gonzlez, F., Mller, A., Dendooven, P., Enghardt, W., Fiedler, F., Kormoll, T., Roemer, K., Petzoldt, J., Wagner, A. & Pausch, G. (2014). Range assessment in particle therapy based on prompt-ray timing measurements, *Phys Med Biol* **59**: 5399.

Goo, H. & Goo, J. M. (2017). Dual-energy ct: New horizon in medical imaging, *Korean Journal of Radiology* **18**: 555.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*, MIT Press. `http://www.deeplearningbook.org`.

Griffin, J. (1966). A statistical model of intermediate structure, *Physical Review Letters (U.S.)* **Vol: 17**.

Hall, E. J. (2000). adiobiology of the radiologist, 5th ed.

Hayakawa, Y., T. J. A. N. H. K. S. M. W. T. T. H. & Tsujii, H. (1995). Acoustic pulse generated in a patient during treatment by pulsed proton radiation beam, *Radiation Oncology Investigations* **3**: 42 – 45.

Helmbrecht, S., Enghardt, W., Parodi, K., Didinger, B., Debus, J., Kunath, D., Priegnitz, M. & Fiedler, F. (2013). Analysis of metabolic washout of positron emitters produced during carbon ion head and neck radiotherapy, *Med Phys* **40**(9): 091918.

Hernandez, L. M. (2017). *Low-dose ion-based transmission radiography and tomography for optimization of carbon ion-beam therapy*, phd thesis, Ludwig-Maximilians-University, Muenchen.

Highland, V. (1975). Some practical remarks on multiple scattering, *Nuclear Instruments and Methods* **129**: 497–499.

Hinton, G. & Sejnowski, T. (1999). *Unsupervised learning foundations of neural computation*, Vol. 38.

Hotelling, H. (1933). Relations between two sets of variates, *Biometrika* **28**: 321–377.

Huang, Z. (2020). *Characterization and experimental validation of a novel prompt gamma prediction algorithm for range verification of proton therapy*, phd thesis, Ludwig-Maximilians-University, Muenchen.

Hueso-Gonzalez, F., Enghardt, W., Fiedler, F., Golnik, C., Janssens, G., Petzoldt, J., Prieels, D., Priegnitz, M., Roemer, K., Smeets, J., Vander Stappen, F., Wagner, A. & Pausch, G. (2015). First test of the prompt gamma ray timing method with heterogeneous targets at a clinical proton therapy facility, *Phys Med Biol* **60**: 6247–6272.

Hueso-Gonzalez, F., Rabe, M., Ruggieri, T., Bortfeld, T. & Verburg, J. (2018). A full-scale clinical prototype for proton range verification using prompt gamma-ray spectroscopy, *Phys Med Biol* **63**.

ICRU90 (2016). Icru report 90, key data for ionizing-radiation dosimetry: Measurement standards and applications, *International Commission on Radiation Units & Measurements* **14**(1).

Inaniwa, T., Kanematsu, N., Furukawa, T. & Hasegawa, A. (2011). A robust algorithm of intensity modulated proton therapy for critical tissue sparing and target coverage, *Phys Med Biol* **56**: 4749–70.

Ivanchenko, V., Incerti, S., Allison, J., Bagulya, A., Brown, J., Champion, C., Elles, S., Francis, Z., Grichine, V., Ivantchenko, A., Jacquemier, J., Karamitros, M., Maire, M., Mantero, A., Pandola, L., Raine, M., Reis, M., Santin, G., Sawkey, D. & Yamashita, T. (2014). Geant4 electromagnetic physics: improving simulation performance and accuracy, p. 03101.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer.

Janssen, F. M., Landry, G., Cambraia Lopes, P., Dedes, G., Smeets, J., Schaart, D. R., Parodi, K. & Verhaegen, F. (2014). Factors influencing the accuracy of beam range estimation in proton therapy using prompt gamma emission, *Phys Med Biol* **59**(15): 4427–41.

Janssens, G., de Xivry, J. O., Fekkes, S., Dekker, A., Macq, B., Lambin, P. & van Elmpt, W. (2009). Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy, *Med Phys* **36**(9): 4268–76.

Janssens, G., Jacques, L., Xivry, J., Geets, X. & Macq, B. (2011). Diffeomorphic registration of images with variable contrast enhancement, *International journal of biomedical imaging* **2011**: 891585.

Jeffers, J. & Reinders, J. (2013). Intel xeon phi coprocessor high-performance programming, *Newnes* .

Jones, D. (1994). Icru report 50prescribing, recording and reporting photon beam therapy, *International Commission on Radiation Units & Measurements* **21**: 833–834.

Jones, K., Vander Stappen, F., Bawiec, C., Janssens, G., Lewin, P., Prieels, D., Solberg, T., Sehgal, C. & Avery, S. (2015). Experimental observation of acoustic emissions generated by a pulsed proton beam from a hospital-based clinical cyclotron, *Med Phys* **42**: 7090–7097.

Kamada, K., Yanagida, T., Pejchal, J., Nikl, M., Endo, T., Tsutsumi, K., Fujimoto, Y., Fukabori, A. & Yoshikawa, A. (2012). Crystal growth and scintillation properties of ce doped $gd_3(ga, al)_5o_{12}$ single crystals, *IEEE Transactions on Nuclear Science* **59**(5): 2112–2115.

Kang, B. H. & Kim, J. W. (2009). Monte Carlo design study of a gamma detector system to locate distal dose falloff in proton therapy, *IEEE Transactions on Nuclear Science* **56**(1): 46–50.

Kelleter, L., Wroska, A., Besuglow, J., Konefa, A., Laihem, K., Leidner, J., Magiera, A., Parodi, K., Rusiecka, K., Stahl, A. & Tessonnier, T. (2017). Spectroscopic study of prompt-gamma emission for range verification in proton therapy, *Physica Medica* **34**.

Kellnberger, S., Assmann, W., Lehrack, S., Reinhardt, S., Thirolf, P., Queirs, D., Sergiadis, G., Dollinger, G., Parodi, K. & Ntziachristos, V. (2016). Ionoacoustic tomography of the proton bragg peak in combination with ultrasound and optoacoustic imaging, *Scientific reports* **6**: 29305.

Kim, J.-W. (2009). Pinhole camera measurements of prompt gamma-rays for detection of beam range variation in proton therapy, *Journal of The Korean Physical Society - J KOREAN PHYS SOC* **55**.

Knopf, A. C. & Lomax, A. (2013). In vivo proton range verification: a review, *Phys Med Biol* **58**(15): R131–60.

Knopf, A., Parodi, K., Bortfeld, T., Shih, H. A. & Paganetti, H. (2009). Systematic analysis of biological and physical limitations of proton beam range verification with offline pet/ct scans, *Phys Med Biol* **54**(14): 4477–95.

Kozlovsky, B., Murphy, R. & Ramaty, a. (2002). Nuclear deexcitation gamma-ray lines from accelerated particle interactions, *The Astrophysical Journal Supplement Series* **141**: 523.

Krimmer, J., Angellier, G., Balleyguier, L., Dauvergne, D., Freud, N., Herault, J., Ltang, J., Mathez, H., Pinto, M., Testa, E. & Zoccarato, Y. (2017). A cost-effective monitoring technique in particle therapy via uncollimated prompt gamma peak integration, *Applied Physics Letters* **110**: 154102.

Krimmer, J., Chevallier, M., Constanzo, J., Dauvergne, D., Rydt, D., Dedes, G., Freud, N., Henriquet, P., La Tessa, C., Ltang, J., Pleskac, R., Pinto, M., Ray, C., Reithinger, V., Richard, M., Rinaldi, I., Roellinghoff, F., Schuy, C., Testa, E. & Testa, M. (2015). Collimated prompt gamma tof measurements with multi-slit multi-detector configurations, *Journal of Instrumentation* **10**: P01011.

Krimmer, J., Dauvergne, D., Letang, J. M. & Testa, E. (2018). Prompt-gamma monitoring in hadrontherapy: A review, *Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment* **878**: 58–73.

Kubota, K. (2001). From tumor biology to clinical pet: a review of positron emission tomography (pet) in oncology, *Ann Nucl Med* **15**(6): 471–86.

Kurosawa, S., Kubo, H., Ueno, K., Kabuki, S., Iwaki, S., Takahashi, M., Taniue, K., Higashi, N., Miuchi, K., Tanimori, T., Kim, D. & Kim, J. (2012). Prompt gamma detection for range verification in proton therapy, *Current Applied Physics - CURR APPL PHYS* **12**.

Kurz, C., Dedes, G., Resch, A., Reiner, M., Ganswindt, U., Nijhuis, R., Thieke, C., Belka, C., Parodi, K. & Landry, G. (2015). Comparing cone-beam ct intensity correction methods for dose recalculation in adaptive intensity-modulated photon and proton therapy for head and neck cancer, *Acta Oncol* **54**(9): 1651–7.

Laramore, G. E., Liao, J. J., Rockhill, J. K. & Komarnicky-Kocher, L. T. (2013). *Relative Biological Effectiveness (RBE)*, Springer Berlin Heidelberg, pp. 748–748.

Lehrack, S., Assmann, W., Bertrand, D., Henrotin, S., Herault, J., Heymans, V., Vander Stappen, F., Thirolf, P., Vidal, M., Van de Walle, J. & Parodi, K. (2017). Submillimeter ionoacoustic range determination for protons in water at a clinical synchrocyclotron, *Phys Med Biol* **62**.

Levin, W. P., Kooy, H., Loeffler, J. S. & DeLaney, T. F. (2005). Proton beam therapy, *Br J Cancer* **93**(8): 849–54.

Litzenberg, D., Bajema, J. & Becchetti Jr, F. (1993). On-line monitoring and p.e.t. imaging of proton radiotherapy beams, *IEEE Transactions on Nuclear Science* **40**: 954–956.

Litzenberg, D., Roberts, D., Lee, M., Pham, K., Molen, A., Ronningen & Becchetti Jr, F. (1999). On-line monitoring of radiotherapy beams: Experimental results with proton beams, *Med Phys* **26**: 992–1006.

Llosa, G., Trovato, M., Barrio, J., Etxebeste, A., Muoz, E., Lacasta, C., Oliver, J., Rafecas, M., Solaz, C. & Solevi, P. (2016). First images of a three-layer compton telescope prototype for treatment monitoring in hadron therapy, *Frontiers in Oncology* **6**.

Lozano, I., Dedes, G., Aldawood, S., Liprandi, S., Miani, A., Zoglauer, A., Lauber, K., Thirolf, P. & Parodi, K. (2016). A compton camera prototype simulation study: Camera performance and first tests of range monitoring capabilities, pp. 1–5.

Maccabee, H., Madhvanath, U. & Raju, M. (1969). Tissue activation studies with alpha-particle beams, *Phys Med Biol* **14**: 213–24.

Malmer, C. (2001). Icru report 63. nuclear data for neutron and proton radiotherapy and for radiation protection, *International Commission on Radiation Units & Measurements* **28**.

Matanoski, G. M., Boice, J. D., J., Brown, S. L., Gilbert, E. S., Puskin, J. S. & O'Toole, T. (2001). Radiation exposure and cancer: case study, *Am J Epidemiol* **154**(12 Suppl): S91–8.

McCleskey, M., Kaye, W., Mackin, D., Beddar, S., He, Z. & Polf, J. (2015). Evaluation of a multistage cdznte compton camera for prompt imaging for proton therapy, *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment* **785**.

Min, C. H., , Zhu, X., Grogg, K., El Fakhri, G., Winey, B. & Paganetti, H. (2014). A recommendation on how to analyze in-room PET for in vivo proton range verification using a distal PET surface method, *Technol Cancer Res Treat* **14**: 320–5.

Min, C. H., Kim, C. H., Youn, M. Y. & Kim, J. W. (2006). Prompt gamma measurements for locating the dose falloff region in the proton therapy, *Applied Physics Letters* **89**: 18.

Min, C. H., Lee, H. R., Kim, C. H. & Lee, S. B. (2012). Development of array-type prompt gamma measurement system for in vivo range verification in proton therapy, *Med Phys* **39**(4): 2100–7.

Miyatake, A., Nishio, T. & Ogino, T. (2011). Development of activity pencil beam algorithm using measured distribution data of positron emitter nuclei generated by proton irradiation of targets containing (12)c, (16)o, and (40)ca nuclei in preparation of clinical application, *Med Phys* **38(10)**: 58185829.

Moteabbed, M., Espana, S. & Paganetti, H. (2011). Monte Carlo patient study on the comparison of prompt gamma and PET imaging for range verification in proton therapy, *Phys Med Biol* **56**(4): 1063–82.

Oelfke, U., Lam, G. & Atkins, M. (1996). Proton dose monitoring with pet: Quantitative studies in lucite, *Phys Med Biol* **41**: 177–96.

Owen, H., Lomax, A. & Jolly, S. (2015). Current and future accelerator technologies for charged particle therapy, *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment* **809**.

Paans, A. & Schippers, J. (1993). Proton therapy in combination with PET as monitor: a feasibilitystudy, Vol. 40, pp. 957–959 vol.2.

Paganetti, H. (2012). Range uncertainties in proton therapy and the role of Monte Carlo simulations, *Phys Med Biol* **57**(11): R99–117.

Paganetti, H. (2018). Proton therapy physics.

Park, J., Kim, S., Ku, Y., Kim, C. H., Lee, H., Jeong, H., Lee, S. & Shin, D. (2019). Multi-slit prompt-gamma camera for locating of distal dose falloff in proton therapy, *Nuclear Engineering and Technology* **51**.

Parodi, K. (2012). PET monitoring of hadrontherapy, *Nuclear Medicine Review* **15**: C37–C42.

Parodi, K. & Assmann, W. (2015). Ionoacoustics: A new direct method for range verification, *Modern Physics Letters A* **30**: 1540025.

Parodi, K. & Bortfeld, T. (2006). A filtering approach based on gaussian-powerlaw convolutions for local PET verification of proton radiotherapy, *Phys Med Biol* **51**: 1991–2009.

Parodi, K. & Enghardt, W. (2000). Potential application of PET in quality assurance of proton therapy, *Phys Med Biol* **45**: N151–6.

Parodi, K., Ferrari, A., Sommerer, F. & Paganetti, H. (2007a). Clinical ct-based calculations of dose and positron emitter distributions in proton therapy using the fluka Monte Carlo code, *Phys Med Biol* **52**: 3369–87.

Parodi, K., Paganetti, H., Shih, H. A., Michaud, S., Loeffler, J. S., DeLaney, T. F., Liebsch, N. J., Munzenrider, J. E., Fischman, A. J., Knopf, A. & Bortfeld, T. (2007b). Patient study of in vivo verification of beam delivery and range, using positron emission tomography and computed tomography imaging after proton therapy, *Int J Radiat Oncol Biol Phys* **68**(3): 920–34.

Pascanu, R., Mikolov, T. & Bengio, Y. (2012). Understanding the exploding gradient problem, *ArXiv* **abs/1211.5063**.

Patch, S., Kireeff Covo, M., Jackson, A., Qadadha, Y., Campbell, K., Albright, R., Bloemhard, P., Donoghue, A., Siero, C., Gimpel, T., Small, S., Ninemire, B., Johnson, M. & Phair, L. (2016). Thermoacoustic range verification using a clinical ultrasound array provides perfectly co-registered overlay of the bragg peak onto an ultrasound image, *Phys Med Biol* **61**: 5621–5638.

Pearson, K. (1901). On lines and planes of closest fit to points in space, *Philosophical Magazine* **2**: 559–572.

Peloso, R., Busca, P., Fiorini, C., Abba, A., Geraci, A., Manenti, A., Longoni, A., Padovini, G., Bianchi, C., Poli, G., Erlandsson, K., Hutton, B., Lechner, P., Soltau, H., Strder, L., Pedretti, A., Mullekom, P. & Pallaro, L. (2010). The hicam gamma camera, pp. 1957 – 1960.

Perali, I., Celani, A., Bombelli, L., Fiorini, C., Camera, F., Clementel, E., Henrotin, S., Janssens, G., Prieels, D., Roellinghoff, F., Smeets, J., Stichelbaut, F. & Vander Stappen, F. (2014). Prompt gamma imaging of proton pencil beams at clinical dose rate, *Phys Med Biol* **59**(19): 5849–71.

Phillips, G. (1995). Gamma-ray imaging with compton cameras, *Nuclear Instruments & Methods in Physics Research Section B-beam Interactions With Materials and Atoms - NUCL INSTRUM METH PHYS RES B* **99**: 674–677.

Pinto, M., Bajard, M., Brons, S., Chevallier, M., Dauvergne, D., Dedes, G., Rydt, M., Freud, N., Krimmer, J., La Tessa, C., Ltang, J., Parodi, K., Pleskac, R., Prieels, D., Ray, C., Rinaldi, I., Roellinghoff, F., Schardt, D., Testa, E. & Testa, M. (2015). Absolute prompt-gamma yield measurements for ion beam therapy monitoring, *Phys Med Biol* **60**: 565–594.

Pinto, M., Dauvergne, D., Freud, N., Krimmer, J., Ltang, J., Ray, C., Roellinghoff, F. & Testa, E. (2014). Design optimization of a tof-based collimated camera prototype for online hadrontherapy monitoring, *Phys Med Biol* **59**: 7653.

Polf, J. & Parodi, K. (2015). Imaging particle beams for cancer treatment, *Physics Today* **68**: 28–33.

Polf, J., Peterson, S., Ciangaru, G., Gillin, M. & Beddar, S. (2009a). Prompt gamma-ray emission from biological tissues during proton irradiation: A preliminary study, *Phys Med Biol* **54**: 731–43.

Polf, J., Peterson, S., McCleskey, M., Roeder, B., Spiridon, A., Beddar, S. & Trache, L. (2009b). Measurement and calculation of characteristic prompt gamma ray spectra emitted during proton irradiation, *Phys Med Biol* **54**: N519–27.

Priegnitz, M., Helmbrecht, S., Janssens, G., Perali, I., Smeets, J., Vander Stappen, F., Sterpin, E. & Fiedler, F. (2015). Measurement of prompt gamma profiles in inhomogeneous targets with a knife-edge slit camera during proton irradiation, *Phys Med Biol* **60**(12): 4849–71.

PTCOG (2020). Facilities in operation. https://www.ptcog.ch/index.php/facilities-in-operation.

Rahman, R. (2013). *Intel Xeon Phi Coprocessor Architecture and Tools*.

Rehman, J., Syed, Z., Ahmad, N., Khalid, M., Khan, M., Gilani, Z., Ullah, I., Nasar, G., Akhtar, M. M. & Usmani, M. (2018). Intensity modulated radiation therapy: A review of current practice and future outlooks, *Journal of Radiation Research and Applied Sciences* **11**.

Remmele, S., Hesser, J., Paganetti, H. & Bortfeld, T. (2011). A deconvolution approach for pet-based dose reconstruction in proton therapy, *Phys Med Biol* **56**: 7601–19.

Resch, A. F., Landry, G., Kamp, F., Cabal, G., Belka, C., Wilkens, J. J., Parodi, K. & Dedes, G. (2017). Quantification of the uncertainties of a biological model and their impact on variable rbe proton treatment plan optimization, *Phys Med* **36**: 91–102.

Ribansky, I., Oblozinsky, P. & Betak, E. (1973). Pre-equilibrium decay and the exciton model, *Nuclear Physics A* **205**.

Richard, M. H., Chevallier, M., Dauvergne, D., Freud, N., Henriquet, P., Le Foulher, F., Letang, J. M., Montarou, G., Ray, C., Roellinghoff, F., Testa, E., Testa, M. & Walenta, A. H. (2011). Design guidelines for a double scattering compton camera for prompt-gamma imaging during ion beam therapy: A Monte Carlo simulation study, *IEEE Transactions on Nuclear Science* **58**(1): 87–94.

Richter, C., Pausch, G., Barczyk, S., Priegnitz, M., Keitz, I., Thiele, J., Smeets, J., Stappen, F. V., Bombelli, L., Fiorini, C., Hotoiu, L., Perali, I., Prieels, D., Enghardt, W. & Baumann, M. (2016). First clinical application of a prompt gamma based in vivo proton range verification system, *Radiotherapy Oncology* **118**(2): 232–7.

Roellinghoff, F., Benilov, A., Dauvergne, D., Dedes, G., Freud, N., Janssens, G., Krimmer, J., Ltang, J., Pinto, M., Prieels, D., Ray, C., Smeets, J., Stichelbaut, F. & Testa, E. (2014). Real-time proton beam range monitoring by means of prompt-gamma detection with a collimated camera, *Phys Med Biol* **59**: 1327–38.

Ruder, S. (2016). *An overview of gradient descent optimization algorithms*.

Salvat, F. (2013). A generic algorithm for Monte Carlo simulation of proton transport, *Nuclear Instruments and Methods in Physics Research B* **316**: 144–159.

Salvat, F, F.-V. J. M. B. J. & Sempau, J. (1996). Penelope, and algorithm and computer code for Monte Carlo simulation of electron-photon showers., *Spain: N. p.* .

Salvat, F., Jose, M. F.-V. & Josep, S. (2011). Penelope-2011: A code systemfor Monte Carlo simulation ofelectron and photon transport.

Sawakuchi, G., Titt, U., Mirkovic, D. & Mohan, R. (2008). Density heterogeneities and the influence of multiple coulomb and nuclear scatterings on the bragg peak distal edge of proton therapy beams, *Phys Med Biol* **53**: 4605–19.

Schaffner, B. & Pedroni, E. (1998). The precision of proton range calculations in proton radiotherapy treatment planning: Experimental verification of the relation between ct-hu and proton stopping power, *Phys Med Biol* **43**: 1579–92.

Schell, S. & Wilkens, J. J. (2010). Advanced treatment planning methods for efficient radiation therapy with laser accelerated proton and ion beams, *Med Phys* **37**(10): 5330–40.

Schlegel, W. (2006). *New Technologies in 3D Conformal Radiation Therapy: Introduction and Overview*, pp. 1–6.

Schmid, S., Landry, G., Thieke, C., Verhaegen, F., Ganswindt, U., Belka, C., Parodi, K. & Dedes, G. (2015). Monte Carlo study on the sensitivity of prompt gamma imaging to proton range variations due to interfractional changes in prostate cancer patients, *Phys Med Biol* **60**(24): 9329–47.

Schneider, W., Bortfeld, T. & Schlegel, W. (2000). Correlation between ct numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions, *Phys Med Biol* **45**: 459–78.

Schulz, K., Ulerich, R., Malaya, N., Bauman, P., Stogner, R. & Simmons, C. (2012). Early experiences porting scientific applications to the many integrated core (mic) platform.

Sempau, J., Sanchez-Reyes, A., Salvat, F., Tahar, H., Jiang, S. & Fernndez-Varea, J. (2001). Monte Carlo simulation of electron beams from an accelerator head using penelope, *Phys Med Biol* **46**: 1163–86.

Smeets, J., Roellinghoff, F., Prieels, D., Stichelbaut, F., Benilov, A., Busca, P., Fiorini, C., Peloso, R., Basilavecchia, M., Frizzi, T., Dehaes, J. C. & Dubus, A. (2012). Prompt gamma imaging with a slit camera for real-time range control in proton therapy, *Phys Med Biol* **57**(11): 3371–405.

Smith, R., Jakas, M., Ashworth, D., Oven, B., Bowyer, M., Chakarov, I. & Webb, R. (1997). Atomic and ion collisions in solids and at surfaces: Theory, simulation, and applications, *Cambridge University Press* .

Sorge, H., Stoecker, H. & Greiner, W. (1989). Relativistic quantum molecular dynamics approach to nuclear collisions at ultrarelativistic energies, *Nuclear Physics A* **498**: 567–576.

Souris, K., Lee, J. & Sterpin, E. (2016). Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core cpu architectures, *Med Phys* **43**: 1700–1712.

Sterpin, E., Janssens, G., Smeets, J., Stappen, F., Prieels, D., Priegnitz, M., Perali, I. & Vynckier, S. (2015). Analytical computation of prompt gamma ray emission and detection for proton range verification, *Phys Med Biol* **60**: 4915.

Sterpin, E., Sorriaux, J., Souris, K., Vynckier, S. & Bouchard, H. (2014). A fano cavity test for Monte Carlo proton transport algorithms, *Med Phys* **41**: 011706.

Sterpin, E., Sorriaux, J. & Vynckier, S. (2013). Extension of penelope to protons: Simulation of nuclear reactions and benchmark with geant4, *Med Phys* **40**: 111705.

Stichelbaut & Jongen (2003). Verification of the proton beam position in the patient by the detection of prompt gamma-rays emission, *39th Meeting of particle Therapy Co-Operative Group* pp. 26–29.

Sulak, L., Armstrong, T., Baranger, H., Bregman, M., Levi, M., Mael, D., Strait, J., Bowen, T., Pifer, A., Polakos, P., Bradner, H., Parvulescu, A., Jones, W. & Learned, J. (1979). Experimental studies of the acoustic signature of proton beams traversing fluid media, *Nuclear Instruments and Methods* **161**: 203–217.

Teoh, M., Clark, C. H., Wood, K., Whitaker, S. & Nisbet, A. (2011). Volumetric modulated arc therapy: a review of current literature and clinical use in practice, *Br J Radiol* **84**(1007): 967–96.

Testa, M., Verburg, J., Rose, M., Min, C. H., Tang, S., Bentefour, E. H., Paganetti, H. & Lu, H.-M. (2013). Proton radiography and proton computed tomography based on time-resolved dose measurements, *Phys Med Biol* **58**: 8215–8233.

Tian, L., Landry, G., Dedes, G., Kamp, F., Pinto, M., Belka, C. & Parodi, K. (2020). A new treatment planning approach accounting for prompt gamma range verification and interfractional anatomical changes, *Phys Med Biol* **65**: 095005.

Tian, L., Landry, G., Dedes, G., Kamp, F., Pinto, M., Niepel, K., Belka, C. & Parodi, K. (2018). Toward a new treatment planning approach accounting for in vivo proton range verification, *Phys Med Biol* **63**: 215025.

Tilley, D., Weller, H. & Cheves, C. (1993). Energy levels of light nuclei a = 16-17, *Nuclear Physics A - NUCL PHYS A* **564**: 1–183.

Tobias, C. A., Lawrence, J. H., Born, J. L., Mccombs, R. K., Roberts, J. E., Anger, H. O., Lowbeer, B. V. A. & Huggins, C. B. (1958). Pituitary irradiation with high-energy proton beams - preliminary report, *Cancer Research* **18**(2): 121–&.

Todd, W., Nightingal, J. & Everett, D. (1974). A proposed gamma camera, *Nature* p. 251:132.

Valverde, A., Brodeur, M., Ahn, T., Allen, J., Bardayan, D., Becchetti Jr, F., Blankstein, D., Brown, G., Burdette, D., Frentz, B., Gilardy, G., Hall, M., King, S., Kolata, J., Long, J., Macon, K., Nelson, A., O'Malley, P., Skulski, M. & Vande Kolk, B. (2018). Precision half-life measurement of c 11: The most precise mirror transition ft value, *Physical Review C* **97**.

Venkatesh, E. & Elluru, S. (2017). Cone beam computed tomography: basics and applications in dentistry, *Journal of Istanbul University Faculty of Dentistry* **51**.

Verburg, J. (2015). Reducing range uncertainty in proton therapy, *PhD dissertation, Technische Universiteit Eindhoven* .

Verburg, J. M., Riley, K., Bortfeld, T. & Seco, J. (2013). Energy- and time-resolved detection of prompt gamma-rays for proton range verification, *Phys Med Biol* **58**(20): L37–49.

Verburg, J., Shih, H. & Seco, J. (2012). Simulation of prompt gamma-ray emission during proton radiotherapy, *Phys Med Biol* **57**: 5459–72.

Vynckier, S., Derreumaux, S., Richard, F., Bol, A., Michel, C. & Wambersie, A. (1993). Is it possible to verify directly a proton-treated plan using positron emission tomography?, *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* **26**: 275–7.

Wang, E., Zhang, Q., Shen, B., Zhang, G., Lu, X., Wu, Q. & Wang, Y. (2014). High-performance computing on the intel xeon phi.

Weisskopf, V. (1937). Statistics and nuclear reactions, *Physical Review - PHYS REV X* **52**: 295–303.

Werner, T., Berthold, J., Hueso-Gonzlez, F., Kgler, T., Petzoldt, J., Roemer, K., Richter, C., Rinscheid, A., Straessner, A., Enghardt, W. & Pausch, G. (2019). Processing of prompt gamma-ray timing data for proton range measurements at a clinical beam delivery, *Phys Med Biol* **64**.

Wilson, R. (1946). Radiological use of fast protons, *Radiology* **47(5)**: 487–491.

Xie, Y. H., Bentefour, E., Janssens, G., Smeets, J., Vander Stappen, F., Hotoiu, L., Yin, L. S., Dolney, D., Avery, S., O'Grady, F., Prieels, D., McDonough, J., Solberg, T. D., Lustig, R. A., Lin, A. & Teo, B. K. K. (2017). Prompt gamma imaging for in vivo

range verification of pencil beam scanning proton therapy, *International Journal of Radiation Oncology Biology Physics* **99**(1): 035019.

Xu, D. & He, Z. (2007). Gamma-ray energy-imaging integrated spectral deconvolution, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **574**(1): 98 – 109.

Yang, Y. & Bednarz, B. (2013). Consistency evaluation between egsnrc and geant4 charged particle transport in an equilibrium magnetic field, *Phys Med Biol* **58**: N47–N58.

Zhu, X. & Fakhri, G. (2013). Proton therapy verification with PET imaging, *Theranostics* **3**: 731–740.

Ziegler, J., Littmark, U. & Biersack, J. (1985). The stopping and range of ions in solids, *Pergamon New York* .